MECHANICAL ENGINEERING SERIES

Giancarlo Genta
Lorenzo Morello

The Automotive Chassis

Volume 1: Components Design

Springer

The Automotive Chassis

www.TechnicalBooksPDF.com

Mechanical Engineering Series

Frederick F. Ling
Editor-in-Chief

The Mechanical Engineering Series features graduate texts and research monographs to address the need for information in contemporary mechanical engineering, including areas of concentration of applied mechanics, biomechanics, computational mechanics, dynamical systems and control, energetics, mechanics of materials, processing, production systems, thermal science, and tribology.

Advisory Board/Series Editors

Applied Mechanics	F.A. Leckie University of California, Santa Barbara
	D. Gross Technical University of Darmstadt
Biomechanics	V.C. Mow Columbia University
Computational Mechanics	H.T. Yang University of California, Santa Barbara
Dynamic Systems and Control/ Mechatronics	D. Bryant University of Texas at Austin
Energetics	J.R. Welty University of Oregon, Eugene
Mechanics of Materials	I. Finnie University of California, Berkeley
Processing	K.K. Wang Cornell University
Production Systems	G.-A. Klutke Texas A&M University
Thermal Science	A.E. Bergles Rensselaer Polytechnic Institute
Tribology	W.O. Winer Georgia Institute of Technology

For other titles published in this series, go to
http://www.springer.com/1161

www.TechnicalBooksPDF.com

Giancarlo Genta • Lorenzo Morello

The Automotive Chassis

Vol. 1: Components Design

Springer

www.TechnicalBooksPDF.com

Prof. Dr. Giancarlo Genta
Politecnico Torino
Dipto. Meccanica
Corso Duca degli Abruzzi, 24
10129 Torino
Italy
giancarlo.genta@polito.it

Prof. Dr. Lorenzo Morello
Politecnico di Torino
Ingegneria dell'Autoveicolo
via Nizza, 230
10126 Torino
Italy
lorenzo.morello@polito.it

ISBN: 978-1-4020-8674-8 e-ISBN: 978-1-4020-8676-2

Library of Congress Control Number: 2008937827

© 2009 Springer Science+Business Media B.V.
No part of this work may be reproduced, stored in a retrieval system, or transmitted in any form or by any means, electronic, mechanical, photocopying, microfilming, recording or otherwise, without written permission from the Publisher, with the exception of any material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work.

Printed on acid-free paper

9 8 7 6 5 4 3 2 1

springer.com

www.TechnicalBooksPDF.com

Contents

ABOUT THE AUTHORS ix

FOREWORD xiii

PREFACE xv

ACKNOWLEDGEMENTS xix

LIST OF SYMBOLS xxi

I WHEELS, STRUCTURES AND MECHANISMS 1

INTRODUCTION TO PART I 3

1 HISTORICAL EVOLUTION 7
- 1.1 Introduction . . . 7
- 1.2 Rigid axle mechanical linkages . . . 9
- 1.3 The independent suspension mechanical linkages . . . 19
- 1.4 Wheels and tires . . . 33
- 1.5 Brakes . . . 41
- 1.6 Chassis frame . . . 45

2 WHEELS AND TIRES 53
- 2.1 Description . . . 53
- 2.2 Tire operation . . . 59

www.TechnicalBooksPDF.com

2.3 Rolling radius 74
2.4 Rolling resistance 75
2.5 Static Forces 88
2.6 Longitudinal Force 90
2.7 Cornering forces 100
2.8 Interaction between longitudinal and side forces 117
2.9 Outline on dynamic behavior 125
2.10 Testing 129

3 SUSPENSIONS 133
3.1 Introduction 133
3.2 Independent suspensions 139
3.3 Semi-independent suspensions 190
3.4 Rigid axle suspensions 195
3.5 Industrial vehicles suspensions 199
3.6 Design and testing 207

4 STEERING SYSTEM 239
4.1 Introduction 239
4.2 Steering mechanism 241
4.3 Rack and pinion steering box 248
4.4 Screw and sector steering box 251
4.5 Steering column 253
4.6 Power steering 255
4.7 Design and testing 261

5 BRAKING SYSTEM 269
5.1 Introduction 269
5.2 Car brakes 272
5.3 Industrial vehicle brakes 290
5.4 Design and testing 297

6 CONTROL SYSTEMS 317
6.1 Steering control 318
6.2 Brake control 326
6.3 Suspension control 340

7 CHASSIS STRUCTURES 351
7.1 Underbody 351
7.2 Subframe 356
7.3 Industrial vehicle frames 360
7.4 Structural tasks 364
7.5 Structural design 371
7.6 Structural testing 378

www.TechnicalBooksPDF.com

II TRANSMISSION DRIVELINE 383

INTRODUCTION TO PART II 385

8 HISTORICAL EVOLUTION 393
8.1 Manual gearbox . . . 395
8.2 Friction clutches . . . 407
8.3 Automatic gearboxes . . . 411

9 MANUAL GEARBOXES 425
9.1 Manual gearbox classification . . . 425
9.2 Mechanical efficiency . . . 428
9.3 Manual automobile gearboxes . . . 430
9.4 Manual gearboxes for industrial vehicles . . . 437

10 SHIFTING MECHANISMS 449
10.1 Internal shifting mechanisms . . . 449
10.2 External shifting mechanisms . . . 453

11 START-UP DEVICES 461
11.1 Friction clutch . . . 461
11.2 Start-up devices for automatic gearboxes . . . 473

12 SYNCHRONIZERS 489
12.1 Description . . . 489
12.2 Design criteria . . . 496

13 DIFFERENTIALS AND FINAL DRIVES 505
13.1 Differentials and final drives . . . 506
13.2 All wheel drive transfer boxes . . . 510
13.3 Outline of differential theory . . . 513
13.4 Types of self-locking differentials . . . 520
13.5 Differential effect on vehicle dynamics . . . 523

14 SHAFTS AND JOINTS 533
14.1 Propeller shafts . . . 534
14.2 Half shafts . . . 536
14.3 Universal joints . . . 538
14.4 Constant speed joints . . . 540

15 AUTOMATIC GEARBOXES 543
15.1 General issues . . . 543
15.2 Car gearboxes with fixed rotation axis . . . 547
15.3 Epicycloidal car gearboxes . . . 554
15.4 Car CVTs . . . 567
15.5 Gearboxes for industrial vehicles . . . 575
15.6 Control strategies . . . 579

www.TechnicalBooksPDF.com

16 DESIGN AND TESTING **593**
16.1 Transmission mission 593
16.2 Gears 597
16.3 Shafts 606
16.4 Bearings 607
16.5 Lubricants 608
16.6 Housings and seals 610
16.7 Outline of test technologies 612

REFERENCES OF VOLUME I **617**

INDEX **621**

www.TechnicalBooksPDF.com

ABOUT THE AUTHORS

Giancarlo Genta

Giancarlo Genta received degrees in aeronautical engineering (1970) and aerospace engineering (1971) at the Politecnico of Turin. He immediately began his career at the Politecnico as Assistant of Machine Design and Technologies.

Dr. Genta has been Visiting Professor of Astronautical Propulsion Systems since 1976 and of Vehicle Mechanics since 1977 and, more recently, of Vehicle System Design in the school of Mechanical Engineering and Automotive Engineering.

He was appointed Associate Professor of Aeronautical Engines Design in 1983, at the Aerospace Engineering School of the Politecnico of Turin, becoming full professor of the same course in 1990.

He was elected Director of the Mechanical Engineering Department of the Politecnico from 1989 to 1995. He has been teaching the course of Applied Stress Analysis II for the Master of Science of the University of Illinois at the Politecnico of Turin.

He has also taught many courses in Italy and abroad as part of development cooperation projects in Kenya (two years), Somalia (six months), India (one month) and at the Bureau International du Travail (Italy).

Dr. Genta has been honorary member of the Academy of Sciences of Turin since 1996, and of the International Academy of Astronautics since 1999; he was elected full member of the same Academy in 2006.

He has coordinated the Research Doctorate in Mechatronics, since 1997.

His research activities, primarily in the field of Machine Design, have focused on static and dynamic structural analysis.

www.TechnicalBooksPDF.com

He has studied the magnetic suspension of rotating parts, vehicle dynamics and related control systems, and was one of the promoters of the Interdepartmental Laboratory on Mechatronics, where he performs research activities on magnetic bearings, moving robots and vehicle mechanics.

Dr. Genta is author of more than 270 scientific publications, covering many aspects of mechanical design, published by Italian, English and American magazines or presented in Congresses.

He has written text books on Vehicle Mechanics (published in Italian and English), adopted as a reference in some Italian and American universities. He has also written monographs on composite material design, on the storage of energy in flywheels (published in English and translated in Russian), on Rotating Systems Dynamics and popular books on space exploration.

Lorenzo Morello

Lorenzo Morello received his degree in Mechanical Automotive Engineering in 1968 at the Politecnico of Turin.

He immediately began his career at the Politecnico as Assistant of Machine Design and Technologies.

Leaving the Politecnico in 1971, went to work at a branch of Fiat dedicated to vehicle studies, one that has been joined to the new Research Centre in 1976. He participates in the development of cars and experimental prototypes for the ESV US Program. He has also developed mathematical models for vehicle suspensions and road holding simulations.

Since 1973 he has been involved in a major project for the development of mathematical models of the vehicle, to address the product policies of the company in facing the first energy crisis; as part of this activity he began the development of a new automatic transmission for reduced fuel consumption and a small direct injection diesel engine to be used on automobiles.

Dr. Morello was appointed manager of the chassis department of the Vehicle Research Unit and has coordinated the development of many research prototypes, such as electric cars, off-road vehicle, trucks and buses.

He was appointed manager of the same Research Unit in 1977 and has been leading a group of about 100 design engineers, dedicated to the development of prototypes. A new urban bus with unitized thin steel sheet body, with spot welded joints, a commercial vehicle that will start production later, a small lightweight urban car, under contract from the National Research Council, and a hybrid car, under contract from the US Department of Energy, were developed in this period of time.

He took responsibility of the Engine Research Unit in 1980; this group, of about 200 people, was primarily dedicated to the development of new car engines. He has managed the development of many petrol engines according to the principle of high turbulence fast combustion. A direct injection diesel engine for cars, many turbocharged pre-chamber diesel engines, a modular two cylinder car engine and many other modified prototypes.

www.TechnicalBooksPDF.com

He was appointed Director of Product Development in 1983; this position includes all applied research activities on Vehicle Products of Fiat Group. The Division included about 400 people, addressed to power train, chassis and bodies studies as well as prototype construction.

Dr. Morello joined Fiat Auto in 1983, to take responsibility for the development of new automotive petrol engines and the direct injection diesel (the first in the world for automobile applications). He was appointed Director for Powertrain Engineering in 1987; the objective of this group was to develop all engines produced by Fiat Auto brands. The most important activity in this period was the development of the new engine family to be produced in Pratola Serra, which included more than 20 different engines.

At the end of his career, he returned to vehicle development in 1994, as director for Vehicle Engineering; this group was addressed to designing and testing bodies, chassis components, electric and electronic systems, wind tunnels, safety center and other facilities.

Dr. Morello retired in 1999 and started a new activity as consultant to the strategic planning of Elasis, a new company in the Fiat Group, entirely dedicated to vehicle applied research.

Along with Fiat Research Center he participated in the planning of courses for the new Faculty on Automotive Engineering of the Politecnico of Turin, and prepared related lecture notes.

He was contract professor of Vehicle System Design and has been contract professor of Automotive Transmission Design for many years at the Politecnico of Turin and the University of Naples; he also published a textbook on this last subject and many articles about the evolution of car technology.

www.TechnicalBooksPDF.com

FOREWORD

These two books on the chassis are part of a series sponsored by ATA (the Italian automotive engineers association) on the subject of automotive engineering; they follow the first book, published in 2005, on automotive transmission.

This event, which I hope will be repeated in the future, is the result of a significant effort lasting more than five years and not yet accomplished.

The Fiat Group is, in fact, well aware of the importance of specialized knowledge on the development and management of a highly competitive product and has turned to the Politecnico of Turin for the opportunity of setting up a course on automotive engineering, addressed to first and second level degree achievement, for specialists who will be dedicated to the development, production and continuous improvement of automotive products.

This course was aimed not only to provide new resources for the company, but also to sustain the company itself in the globalization process, only possible with a cultural homogeneity between parts or services suppliers and people in charge of delocated processes.

This course, operative in Turin since the academic year 1999/2000, has been planned and begun as a result of a project that involved Professors of the Politecnico, addressed to the automotive disciplines and experts of many companies of the Fiat Group; the participation of these experts was not limited to the planning of specialist courses, but was also extended to the preparation of lecture notes and, quite often, to actual teaching activity.

Fiat assigned this task to the Fiat Research Center, for many reasons.

Fiat Research Center (CRF) has the responsibility not only for designing innovative products, but also for developing new processes for product development and production. In addition, CRF must diffuse and make available to the

www.TechnicalBooksPDF.com

company's operating sectors the knowledge that derives from new product development, to assure a quick introduction of competitive products to the market.

Finally, CRF is dedicated not only to automobiles, but also to other automotive products and components and to production systems; for this reason it has been possible to include industrial vehicles and component suppliers, taking for granted a greater emphasis on automobiles.

This task was particularly difficult and involved the participation of many specialists of the Research Center and a number of experts from the operating field; the result of this effort consists not only in an integrated studies plan, but also complete lecture notes and audiovisual aids to support lessons and the activities of students.

The quantity of this material has encouraged us to go further, with the intention of transforming this material into reference books in Italian and, possibly, in the English language.

The Automotive Chassis is dedicated to the design of related components and their integration into the vehicle, in order to obtain customer satisfaction. This book supports the courses of Automotive System Design, Automotive Chassis Design and Automotive Transmission Design that are held at the Politecnico of Turin as part of the Automotive Engineering Course.

ATA, our Italian associations of automotive engineers, has overseen publication of the Italian edition; this task fits well with the institutional objectives of the association, to diffuse and foster automotive culture among young people.

Nevio Di Giusto
CRF and Elasis Chief Executive Officer

www.TechnicalBooksPDF.com

PREFACE

This book is the result of two decades of experience: From one side the experience of teaching courses such as Vehicle Mechanics, Vehicle System Design, Chassis Design and more to students of Engineering; from the other the design praxis of vehicle and chassis components in a large automotive company. This book is addressed primarily to students of automotive engineering and secondarily to all technicians and designers working in this field. It is also addressed to all enthusiasts who are looking for a technical guide.

The tradition and the diversity of disciplines involved in road vehicle design lead us to divide the vehicle into three main subsystems: The engine, the body and the chassis.

The chassis is no longer – as engine and body are – a visible subsystem created in a certain part of the fabrication process; chassis components are assembled, as a matter of fact, directly on the body. For this reason the function of the chassis cannot be assessed separately from the rest of the car.

As we will see better when reading the chapters in this book dedicated to historical evolution, the situation was completely different in the past; in the first cars the *chassis* was defined as a real self-moving subassembly, one that included:

- A structure, usually a ladder *framework*, able to carry on all the remaining components of the vehicle.
- The *suspensions* for the mechanical linkage of wheels with the framework.
- The *wheels* and their tires.

www.TechnicalBooksPDF.com

- The *steering system* for changing wheel angles according to vehicle path.
- The *brake system* for reducing speed or stopping the vehicle.
- The *transmission* for applying engine torque to the driving wheels.

This group of components, after engine assembly, was able to move autonomously, as occurred in many experimental tests, where the body was simulated with a ballast and the chassis moved during the fabrication process from the shop of the car maker to that of the body maker.

Customers often bought a chassis from the car maker to be completed later by a body maker, according to their desire and specification.

In contemporary vehicles this particular architecture and function is only available for industrial vehicles, with the exception of buses where the structure, even if built by a body maker, contributes along with the chassis framework to the total stiffness, forming a kind of unitized body.

In almost every car the chassis structure cannot be separated from the body, being part of its floor (*platform*); sometimes an auxiliary framework is also added to interface the suspension or power train to the body, enabling their pre-assembly at the side of the main assembly line.

Nevertheless tradition and particular technical aspects of these components have justified the development of a particular discipline within vehicle engineering; as a consequence almost all car manufacturers have a technical organization addressed to the chassis, separated from those addressed to body or engine.

Another reason has been added in recent times to justify a different discipline and a specific organization. This is the creation of the so called *technological platforms*: The modern trend of the market calls for an unprecedented product diversification, of a sort never attained in the past; marketing experts sometimes call this phenomenon fragmentation.

This high degree of diversification could not be sustained with acceptable production costs without a strong cross standardization of non-visible or non-specific parts of certain models.

This situation has been well known to all industrial vehicle manufacturers for years. The term 'platform', implying the underbody and front side members, with the addition of the adjective 'technological', describes a set of components substantially equal to the former chassis. The particular technical and scientific issues, the different development cycle, the longer economic life have reinforced the specificity of engineers dedicated to this car subsystem.

The contents of this book are divided into five parts, organized into two volumes.

The first volume describes primary chassis subsystems in two parts.

The first part describes the primary components of the chassis from the tire to the chassis structure, including wheels, suspension, steering and braking systems, not forgetting the control systems now becoming increasingly important because of the diffusion of active and automatic systems.

www.TechnicalBooksPDF.com

The second part is addressed to the transmission and related components; the complexity of this topic justifies a separate presentation.

It should be noticed that, among many car manufacturers, the engineering and production organizations dedicated to this subsystem are integrated into the power train organization, instead of the chassis organization. This has obviously no influence on the technical contents of this book and can be justified by standardization issues and the life cycle of this component, in certain respects more like to the engine than the chassis.

The explanation of chassis components assumes the existence of a general knowledge of mechanical components, which can be gathered through a conventional machine design course. Topics that can be found in a non-specific course are not treated. In the second part in particular gear design will not be approached exhaustively, nor will shaft, bearing and seal design.

Nevertheless, in many parts of this book design and testing information not usually approached in general purpose design courses is introduced and discussed.

We also decided to spend two chapters on the historical evolution of the automotive product; those should enable the reader to appreciate the technical progress of the car in its first hundred years of life. In the opinion of the authors, this subject is useful as technical training and is sometimes inspirational as well.

Only architectures that are typical of the most widely used road vehicles will be considered: cars, with some mention of industrial vehicles. We will not consider other applications, such as motor bicycles, tractors or earth-moving machines and quadricycles.

The second volume is divided into three parts and is entirely addressed to the chassis as a system, considering the contribution of the chassis to vehicle performance, as perceived by the customer and imposed by legislative rules.

The third part is dedicated to an outline of the functions that the vehicle is expected to perform, of customer expectations and pertinent legislation.

In the fourth part the influence of chassis design on vehicle performance is explained. In particular, the longitudinal, transversal and vertical dynamics is explained, along with its influence on speed, acceleration, consumption, breaking capacity and maneuverability (or handling) and comfort.

The fifth part is addressed to mathematical models of the chassis and at large vehicle. Automotive engineers take more and more advantage of mathematical models of virtual prototypes and perform numerical testing before the prototypes are available for physical tests.

Even if mathematical models are based upon calculation codes that are prepared by specialists and are available on the market, we think it necessary to supply students with a clear idea of the methods behind these codes and the approximations these codes imply. The purpose of this section is not to enable specialists to built up their models, but to suggest a correct and responsible usage of their results.

The two books are completed by five appendices.

www.TechnicalBooksPDF.com

The first appendix reviews basic system dynamics, useful for understanding the creation of the mathematical models that are introduced in the fourth and fifth part.

The second appendix is dedicated to two wheel vehicles. The study of two wheel vehicles, in some respects more complicate than four wheel vehicles, is very particular and has nothing to do with cars. In addition, industries that produce motorcycles are in a different category from the car industry.

Nevertheless, there are disciplines common to the two worlds, due to the fact that both vehicles use pneumatic tires as their interface with the ground; knowledge exchange between vehicle engineers in both camps could be of mutual benefit.

The third appendix is dedicated to the particular issues that must be faced when vehicles on wheels will be developed for planets or environments different from the earth. Beginning with the only vehicle of this kind that was developed for the Apollo Project, similarities and differences between conventional vehicles and those that in the future could be utilized for interplanetary exploration will be discussed.

The fourth appendix analyzes various mathematical approaches, sometimes simplified, to interpret the motion of cars after the impact due to an accident.

The last appendix reports the primary data of vehicles of different kinds that are used in explanatory examples in the book; these data also enable students to practice their skills on exercises with a minimum of realism.

Torino, Italy

Giancarlo Genta
Lorenzo Morello

www.TechnicalBooksPDF.com

ACKNOWLEDGEMENTS

The authors wish to thank Fiat Research Center for having made the preparation of these two volumes possible, not only by supporting the cost of this work, but also by supplying a great deal of technical material, essential to produce an updated and application-oriented text.

The authors particularly appreciated the suggestions and information they received from Kamel Bel Knani, Roberto Cappo, Paolo Mario Coeli, Silvio Data, Roberto Puppini and Giuseppe Rovera.

Particular thanks are conveyed to Donatella Biffignandi of the Automobile Museum of Turin for the help and material supplied for the preparation of the historical sections.

The authors are also indebted with their editors: Paul Gilster deeply revised the English text; Natalie Jacobs put the finishing touches on the book.

The first volume of this work has, in addition, benefited from the lecture notes prepared by Fiat Research Center to sustain the teaching activity of the courses of Vehicle System Design, Chassis Design and Automotive Transmission Design, within the course of Automotive Engineering of the Politecnico of Turin and of the Master in Automotive Engineering of the Federico II University of Naples.

The authors' gratitude must also be shown to the companies that supplied part of the material used for the illustrations, mainly in the first volume; in alphabetical order we remember: Audi, Fiat Auto, Getrag, Honda, Iveco, Marelli, Mercedes and Valeo. Without their contribution this book would be neither complete nor topical.

www.TechnicalBooksPDF.com

LIST OF SYMBOLS

a	acceleration; generic distance; distance between center of gravity and front axle
b	generic distance; distance between center of gravity and rear axle
c	viscous damping coefficient; specific heat
d	generic distance, diameter
e	base of natural logarithms
f	rolling coefficient; friction coefficient
f_0	rolling coefficient at zero speed
$\mathbf{f}$	force vector
g	gravity acceleration
h	wheel deflection
h_G	center of gravity height on the ground
k	stiffness
l	wheelbase; length
m	mass
p	pressure
r	radius
s	stopping distance, thickness
t	temperature; time; track
$\mathbf{u}$	displacement vector
v	slipping speed
z	number of teeth
A	area
C	cornering stiffness; damping coefficient
C_γ	camber stiffness

www.TechnicalBooksPDF.com

C_0	cohesiveness
E	energy; Young modulus
F	force
G	shear modulus
H	thermal convection coefficient
I	area moment of inertia
J	quadratic mass moment
K	rolling resistance coefficient; stiffness; thermal conductivity
$\mathbf{K}$	stiffness matrix
M	moment
M_f	braking moment
M_m	engine moment
M_z	self-aligning moment
P	power; tire vertical stiffness; force
P_d	power at the wheel
P_m	power at the engine
P_n	required power
Q	thermal flux
R	undeformed wheel radius; path radius
R_e	rolling radius
R_l	loaded radius
S	surface
T	temperature, force
V	speed; volume
W	weight
α	sideslip angle; road side inclination; angle
α_t	road transverse inclination angle
γ	camber angle
δ	steering angle
ϵ	toe-in, -out; brake efficiency; deformation
η	efficiency
θ	angle; pitch angle
μ	torque transmission ratio; adherence coefficient
μ_p	maximum friction coefficient
μ_x	longitudinal friction coefficient
μ_{x_p}	max longitudinal friction coefficient
μ_{x_s}	slip longitudinal friction coefficient
μ_y	transversal friction coefficient
μ_{y_p}	max transversal friction coefficient
μ_{y_s}	slip transversal friction coefficient
ν	speed transmission ratio; kynematic viscosity
ρ	density
σ	normal pressure; slip
τ	transversal pressure; transmission ratio

www.TechnicalBooksPDF.com

ϕ angle; roll angle, friction angle
ω pulsation; frequency
Φ diameter
Π tire torsion stiffness
χ torsion stiffness
Ω angular speed

www.TechnicalBooksPDF.com

INTRODUCTION TO PART I

The first part of this book is dedicated to the study of the often complex subsystems (such as suspensions) that constitute the chassis. Their primary function is to mediate the exchange of force with the ground, thus obtaining the desired vehicle speed and path.

With reference to the system of coordinates that will be defined in part four, the forces exchanged with the ground can be classified as:

- Forces perpendicular to the ground (vertical for the motion on a plane road); in steady state conditions these forces can be considered constant, but because of obstacles on the road, they are variable; they are a factor in passenger comfort.

- Longitudinal forces; these are primarily due to propulsion (engine and transmission) and braking systems; they are relevant to vehicle speed control.

- Transversal forces; these are due to wheel steering angles, and they are relevant to road holding and stability.

All these forces act on tires that because of their deformable structure, make the dynamic behavior of the vehicle more like a floating or flying vehicle than a rail vehicle.

Although chassis technologies can be defined as mature, we should not underestimate the ongoing evolution of controlled or active systems based upon electronic and informatic technologies.

www.TechnicalBooksPDF.com

In fact the rapid development of automotive electronics, in terms of performance and cost has had and will continue to have a big influence on improving the active safety and comfort of vehicles.

Nor it should be forgotten that in many markets the development and production of chassis systems is being outsourced by car manufacturers to parts manufacturers, who are becoming specialists in their business.

This has been true for many years of brake systems, steering systems and tires; it is now becoming true for suspensions. Given this situation, it is important for those who will address their career to car or parts manufacturing to develop a good understanding of these systems; the development of these components is virtually impossible if separated from that of the vehicle.

As we will see, chassis components have evolved quickly in recent years: today almost all cars feature radial tires with low aspect ratios (the radial dimension is much smaller than transverse dimension) and need suspensions with precise elasto-kinematic behavior. Mc Pherson and double wishbone suspensions share the market as far as front axles are concerned, while a significant percentage of rear axles feature multilink suspensions.

It is unlikely that the kinematic configuration will see new innovations; the same can be said for the steering system, where the wide diffusion of power systems has almost standardized the rack and pinion configuration.

A similar situation can be seen for car brake systems, where disc brakes are widely diffused with the exception of the rear axle of economy cars that preserve the drum solution.

New developments are, however, expected for electronic control systems and the related fields of sensors and actuators, where electromechanical actuators offer more opportunities for performance improvements.

Electronic control systems have initially entered the marked as add-on devices.

The case of the brake antilock system (ABS) is typical: It made significant performance improvements to the brake system at the cost of new and sophisticated components (the electronic control system, wheel speed sensors, a valve group able to regulate the pressure on the brake actuators of the wheel independent of the pedal pressure).

Although the introduction of this system was gradual, it later reached high volumes with consistent cost reduction and now, as a consequence, its diffusion is nearly total. At the same time system performance was improved, offering new possibilities, either in terms of cost reduction (i.e. offering the possibility of incorporating the brake distribution valve function at no cost), or in terms of functions where, with the addition of various sensors vehicle dynamic control has been obtained.

A similar story can be told for power steering systems, initially totally hydraulic; the addition of electronic controls allowed better regulation of the power assistance pressure, reducing the sensitivity of the steering wheel torque to the vehicle speed.

www.TechnicalBooksPDF.com

The present trend consists in substituting an electric electronic system for the hydraulic electronic system: Power assistance comes from a controlled electric motor. This offers the possibility of having an active steering system that can improve vehicle performance while avoiding sudden obstacles.

It is likely that all actuators will become electric in the future, with cost reductions and increased performance; the next step could be to avoid any mechanical linkage between pilot controls (pedal, steering wheel, etc.) and actuators.

This goal has already been reached for the engine, where throttle position or fuel injection quantity are no longer controlled mechanically by the accelerator pedal, but through a *drive-by-wire* system. We can easily foresee for the future a *brake-by-wire* system or a *steer-by-wire* system.

The next step, now a topic of discussion in many technical congresses, is the *corner-by-wire* that is a wheel-suspension group (*corner*) with total electric actuation (driving, braking and steering functions); a system like this could have a significant impact on vehicle performance and architecture.

Similar evolutionary processes are present in the suspension field; a first step is the application of electronic controls to the damping properties of shock absorbers and to the position of the body relative to the ground while the vehicle is standing still (*trim*); this could lead to a suspension where the body position is also controlled dynamically. Such an achievement would simplify the elasto-kinematic requirements of the suspension.

We think that these and other examples offer a view of the possible paths of chassis evolution.

After a chapter dedicated to the historical evolution of the chassis, the most widely used configurations for chassis components will be described. The following components will be considered.

Wheels and tires

Tires will not be studied from the stand point of their product and process design techniques, which are useful in determining their performance. They will be studied almost as a *black-box,* examining their static and dynamic response which is the basis of vehicle static and dynamic response.

A good knowledge of tire performance is fundamental for effective communication between vehicle and tire specialists.

Suspensions

While studying suspensions, the main kinematic schemes will be considered along with their influence on the working angles of the tires, on vehicle roll and pitch. The most important suspension components will be described, such as the primary elastic elements, the secondary elastic elements and the damping elements.

www.TechnicalBooksPDF.com

Steering system

The primary mechanisms of the steering system will be studied along with their mechanical properties; the primary components will be described, such as the steering box and the most important power assistance systems.

Brake system

The most important brake types will be introduced along with their actuation and power assistance systems. Industrial vehicle brake systems will be described separately because they use a different actuation system (pneumatic instead of hydraulic power).

Control systems

As far as chassis control systems are concerned, this volume will describe sensors and actuators in use and the technical target these systems should reach in terms of vehicle dynamics; the most diffused control strategies that the different systems adopt will be also described, while the interaction between control system and dynamic behavior of the vehicle will be examined in the second volume.

Chassis structures

Although this topic could be better tackled in a book dedicated to body design, this chapter will outline the integration of the chassis functions into the body structure and will offer a short description of the primary types of auxiliary frameworks in use on unitized bodies. A short description of industrial vehicle frameworks is also offered.

www.TechnicalBooksPDF.com

1
HISTORICAL EVOLUTION

1.1 INTRODUCTION

It is always challenging to decide whether historical notes should come before or after the description and interpretation of the state of the art.

Arguing for the second alternative is the fact that readers should already have understood the motivations that drive a design decision. In favor of the first alternative, which we have chosen, there is the fact that readers can place the state of the art in context and can better appreciate the effort that has been made by those people who have contributed to it.

The unavoidable drawback of this approach is that some considerations are only superficially introduced in the historical section that are discussed in better detail later; readers will forgive us for repeating some concepts. We suggest that readers interested in chassis history read the historical notes again when they come to the end of the volume.

We begin this chapter with suspensions and steering systems, going on later to describe wheels, tires, brake systems and structures. This emphasis is solely due to the larger impact that suspensions and steering systems have on vehicle architecture and its consequent evolution; steering systems will be described together with suspensions because these two systems are indissoluble from the point of view of designers.

The primary function of a vehicle suspension is to isolate in the best way possible the body, the sprung mass, from the disturbances coming from the uneven surface of the road. To do this, the wheels and the masses integral to them, the so called unsprung masses, are connected to the body with mechanical

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

www.TechnicalBooksPDF.com

linkages that allow their relative motion, primarily in the vertical direction; along this direction forces are transmitted to the body trough elastic and damping elements.

The elastic characteristics of tires also contribute to suspension quality.

Because the path of the wheel motion is not strictly vertical with reference to the car body, but can show components in the other two directions as well, a second function for the suspension can be identified: It should guide the wheel, during its displacement (suspension deflection) in order to avoid undesired motions, in terms of:

- Steering angle of the wheel (*toe-in*, *toe-out*), that can modify vehicle path
- Wheel camber angle, that can affect the cornering stiffness of the tire
- Track variation, which can influence tire longevity
- Wheelbase variation, which can cause resonances with traction and braking force application

These displacements which we will define as secondary, cannot be completely eliminated, but must be accurately designed while defining the kinematic linkages and the stiffness of the reaction points (elasto-kinematics); in fact a design variation of these parameters can affect the vehicle dynamic behavior positively.

This topic will be discussed completely in the suspension chapter.

To reveal in advance what will be discussed later on, we can say that:

- The toe angle of the wheel can correct the vehicle understeering (and over steering) behavior and improve vehicle stability during braking.
- A camber angle equal but opposite to the roll angle, or in other words, a constant perpendicularity of the wheel to the ground allows the maximum exploitation of the tire cornering stiffness.
- An appropriate wheelbase variation can improve the wheel's ability to absorb the effect of obstacles, with positive results for comfort; in addition, the suspension can be designed to have *anti-dive* and *anti-squat* characteristics, minimizing the pitch angle variation as a consequence of the application of driving and braking forces.

It should not be forgotten that four wheels vehicles are, as far as the exchange of forces with the ground is concerned, simply hyperstatic systems; therefore a last function of the suspension is distributing forces exchanged with the ground and making all wheel contact possible also if the road isn't flat.

Many of these considerations, which we will better explain later, were not known at the beginning of the motor era and have been developed quite recently, as compared with the hundred years of life of the automotive product, because of the increase in vehicle speed and improvements in road conditions.

www.TechnicalBooksPDF.com

The first two sections of this chapter trace the evolution of the mechanical linkages of the suspension and of steering system in an attempt to understand, according to documents and drawings of that time, the motivations and ideas of the engineers that developed them.

1.2 RIGID AXLE MECHANICAL LINKAGES

We include under this title the articulated and elastic systems of both suspension and steering mechanism.

The suspension function was already known in the XVI century. Coach bodies were suspended through a leaf spring set fit to the chassis framework, a rigid structure bearing wheel hubs. The free end of the springs was connected to the body through leader belts; Fig. 1.1 offers an interesting example of this kind of configuration in a coach dated about 1650. The steel leaf springs present in this vehicle were introduced during this time; formerly, they were made of wood. There is no component explicitly dedicated to damping suspension oscillations; the internal friction of leaf springs and belts should have been enough to reach the expected level of comfort.

Secondary motions were not present because wheels were rigidly connected; the system was isostatic because of the play of the front steering axle.

FIGURE 1.1. This coach built around 1650 shows the existence of a suspension; the sprung mass includes the passenger compartment only and is connected to the unsprung mass with four leaf springs with leader belts (Automobile Museum of Turin).

www.TechnicalBooksPDF.com

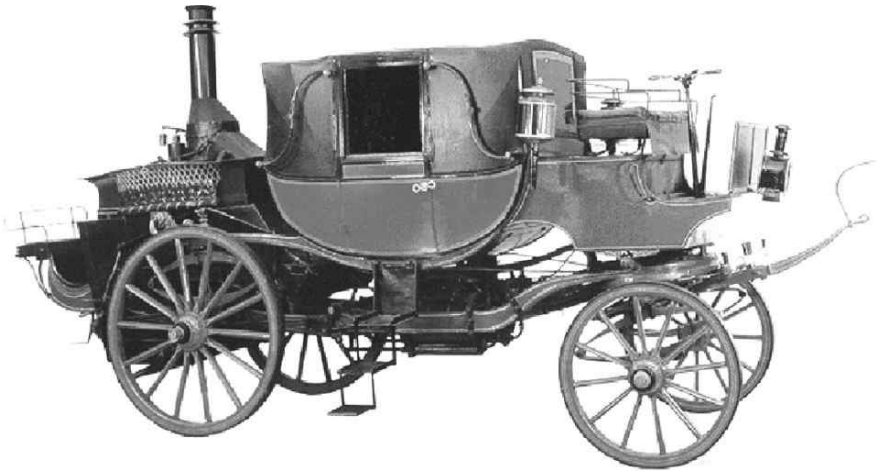


FIGURE 1.2. The Bordino's steam coach was built in 1854. The suspension system is strictly derived from that of a horse carriage. The rear axle is shaped as a crankshaft, where the connecting rods of the two cylinders operate directly (Automobile museum of Turin).

Elliot, an English wheelwright, was credited with the invention of the single rigid axle suspension, using semi elliptical steel leaf springs; in this design the front axle can steer on a pivot in the middle of the axle.

This suspension system, in use on coaches and carriages, was also adopted in the first steam road vehicles, in the XIX century, before the advent of the internal combustion engine. Figure 1.2 shows an example.

The carriage steering system showed the inconveniently reduced the roll-over stability of the vehicle in a turn; when lateral centrifugal forces were applied, the roll-over line (which can be obtained by joining the two contact points with the ground of the wheels on the same side) was shifted closer to the centre of the vehicle.

In 1810 Längensberger envisaged a steering system where the front axle was always parallel to the rear axle but only the front wheel hub was articulated to the axle using a king-pin; the two stub axles were connected through track-rod arms and a track rod, forming an articulated parallelogram. Using this device the vehicle roll over stability was unaffected in turns.

In 1818 a patent was filed in London in the name of Ackermann; this patent described the law the steering angles of the two wheels should follow in order for the wheels to roll correctly, with their symmetry plane containing the local speed vector. This invention found no immediate practical application because there was no real need for carriages to have a steering system different from the turntable steering, and there was no idea about how to satisfy this law in a simple way.

www.TechnicalBooksPDF.com

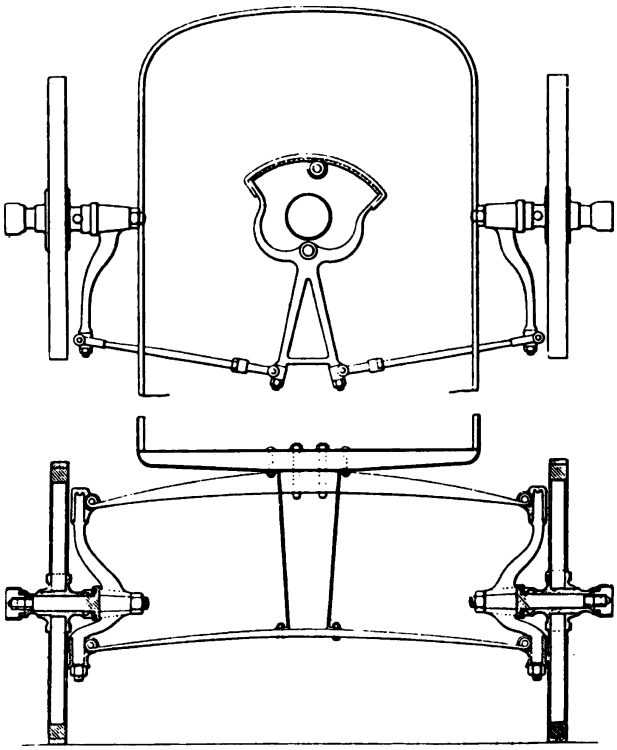


FIGURE 1.3. Drawing of the front steering axle of the Mancelle of Bollée, in 1878; notice the independent double wishbone suspension with transverse double leaf springs.

Jeantaud, another wheelwright, proposed in 1878 a mechanism, perfected from the idea of Längensberger, in which the two track-rod arms were slightly inclined to the middle of the vehicle so that their axes crossed near the middle of the rear axle.

This mechanism obtains with acceptable approximation Ackermann's law and is still in use today on rigid steering axles.

Independent of Jeantaud's idea, the solution presented by Bollée in his Mancelle, also in 1878, was not far away from the Ackermann's law. We like to recall this steering system because it is probably one of the first bound to an automobile: Fig. 1.3 shows the scheme of the front axle; notice that it is an independent suspension with transversal leaf springs, equivalent to the double wishbone mechanism.

We should not assume that independent wheel suspensions were invented in the 1940s, at the end of an era that saw only rigid axles with leaf springs; independent wheel suspensions are sporadically present even in the first cars. Nevertheless, we should assume that the advantages obtained on existing roads and speeds were negligible when compared with the enormous design complications.

www.TechnicalBooksPDF.com

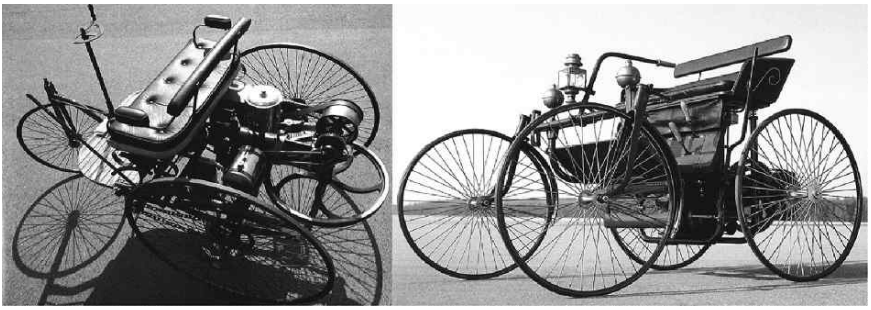


FIGURE 1.4. The first cars with an internal combustion engine, the tricycle of Benz and the *Stahlradwagen* of Daimler give no evidence of great attention to the suspension system. Both lack a front suspension and the second features no suspensions at all, with a suspended seat only.

The first cars with an internal combustion engine, the tricycle of Benz in 1886 and the *Stahlradwagen* (German word for 'steel wheel car') of Daimler in 1889 (see Fig. 1.4) give no evidence of great attention to the suspension system.

The first is in fact a tricycle with a single non-suspended front steering wheel; the second is a four wheel vehicle with no suspension, in which the front axle is balanced on a central horizontal pivot that makes the system isostatic. We can assume that the undoubted genius of these two precursors was focused on developing a reduced weight engine with a significant, but low mass, quantity of energy on board, rather than on passenger comfort.

Let us now examine the most important features of rigid axle suspensions with leaf springs: Axle linkages are in this case integrated with the elastic element.

As far as suspensions are concerned, we should we aware of two typical problems of suspended axles.

- If the engine is part of the sprung mass (and this is usually the case), it is necessary to develop a mechanism to connect the engine shaft or the gearbox output shaft with the wheels, which have a variable relative position.

- Because the steering control (steering wheel or steering bar) must be within reach of the driver, mechanisms must be developed that connect stub axles without affecting the steering angles through the bouncing motion of the suspension.

The solution of these two problems would have preoccupied the first car designers, particularly in terms of the steering system; the multiplicity of solutions in the first cars gives evidence of the difficulty encountered in finding an adequate technical solution to the problem.

www.TechnicalBooksPDF.com

On the Benz car the problem is bypassed, considering that there is no front suspension. On the Daimler, the steering control is fixed to the balancing axle near the pivot point.

As far as the transmission is concerned, on the Benz car a leather belt gearbox compensates for the center line variation of the pulleys with a spring-loaded moving tensioner.

This device is common to other cars of this era and has the advantage of integrating transmission functions and gearbox functions; the gearbox function is obtained by using a number of coupled pulleys with a shiftable belt.

Almost in the same year the idea of using a tooth wheel gearbox was born, in connection with a chain transmission; the Fiat 3 $\frac{1}{2}$ HP of 1899 could be taken as an example of this concept, as shown in the phantom view of Fig. 1.5. The sprocket axis was fixed to the sprung mass, almost in the center of curvature of the suspension path described in the bouncing motion. This design detail allowed the axle to move without affecting the chain length.

This transmission and suspension architecture was widely applied in the first twenty years of automotive history.

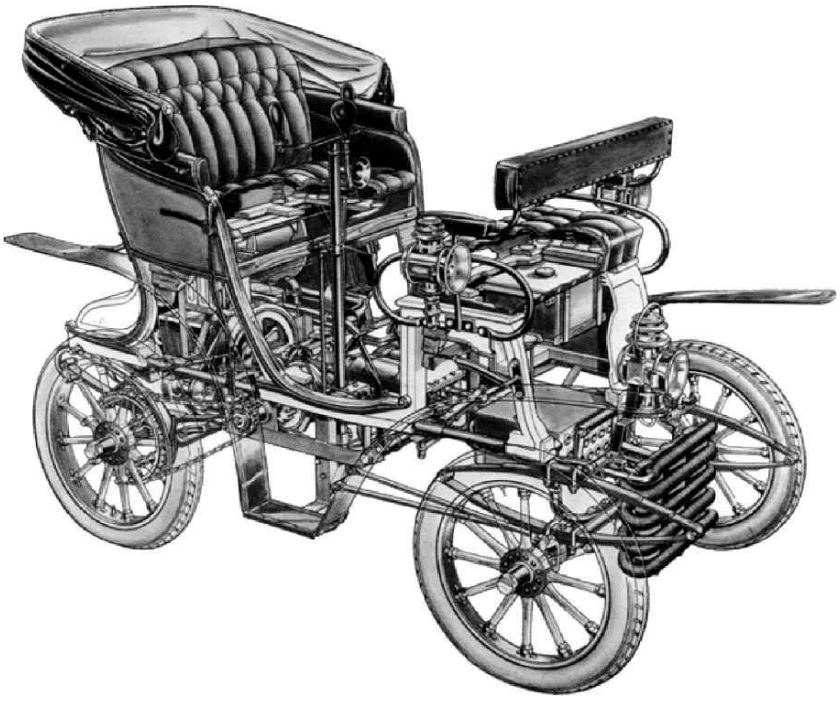


FIGURE 1.5. The Fiat $3\frac{1}{2}$ HP of 1899 features a chain and sprocket transmission, where the sprocket is set in the center of curvature of the path of motion of the rear axle. Notice the Jeantaud steering system.

www.TechnicalBooksPDF.com

On the same car the Jeanteaud steering mechanism connected to the steering bar through a longitudinal linkage can be seen; also notable is that the linkage knuckle mounted on the sprung mass was positioned in the center of curvature of the motion of the axle in its bouncing movement, to avoid undesirable changes of direction while riding on uneven roads.

Almost simultaneously, in 1898, De Dion & Bouton introduced the propeller shaft transmission with universal joints. It is still in use, almost unchanged in its basic elements, in today's cars with front engine and rear drive; it solves optimally any problems connected with suspension motion.

Leaf springs, as we have seen, were already known in the XVII century; at the beginning of the XX century they obtained a satisfactory level of fabrication technology and application know-how. Leaf springs integrate in a single element the elastic function and the structural function; the elasto-kinematic performance of this component was fully adequate for the needs of the cars in that era.

The spring is made with leaves of different length, reduced by constant steps from the top to the bottom; this assembly simulates with good approximation the performance of a uniform stress structure, allowing the maximum value of deformation at a given stress level. This fact is demonstrated in Fig. 1.6, drawn from one of the first manuals of automotive engineering.

The same structure that is flexible on a vertical plane can be quite stiff in a horizontal plane.

The spring was mounted to the chassis with a fixed eye on the mother leaf and with a moving eye articulated to a swinging shackle, or provided with a sliding element; this device allowed the spring to change its length because of the deflection. The fixed eye was on the same side as the fixed point of transmission and steering system, typically behind the front axle and in front of the rear axle.

The suspension architecture could assume many different variants, according to objectives of comfort, cost and weight of the unsprung mass. Leaf springs have been named after their shape, with reference to the ideal elliptical shape, made of two mirror like leaf springs; with reference to Fig. 1.5 front springs are called elliptical, while rear springs are called semi elliptical.

The most commonly used solution featured two semi elliptical springs for each axle; the $\frac{3}{4}$ of ellipse solution (Fig. 1.7, top left) was sometimes adopted on luxury cars, where the side beam of the chassis structure was fit directly to a $\frac{1}{4}$ ellipse spring. This solution allows increased flexibility, though at a higher cost and weight of the unsprung mass.

A three semi elliptical spring solution was also adopted for the rear axle in luxury cars; the function is comparable to that of the $\frac{3}{4}$ ellipse arrangement, where the two $\frac{1}{4}$ ellipse springs are integrated in a single element, fit in the middle to the structure of the chassis. In this case, swinging shackles are made with a universal joint (Fig. 1.7, bottom).

A particular cantilever spring, shown in Fig. 1.7 at the top right, was developed and distributed by Ford. A single semi elliptical spring is fit to a cross beam of the chassis structure, at the mid section; it performs like two $\frac{1}{4}$ ellipse cantilever springs, mounted on the axle; in this case, two swinging shackles must

www.TechnicalBooksPDF.com

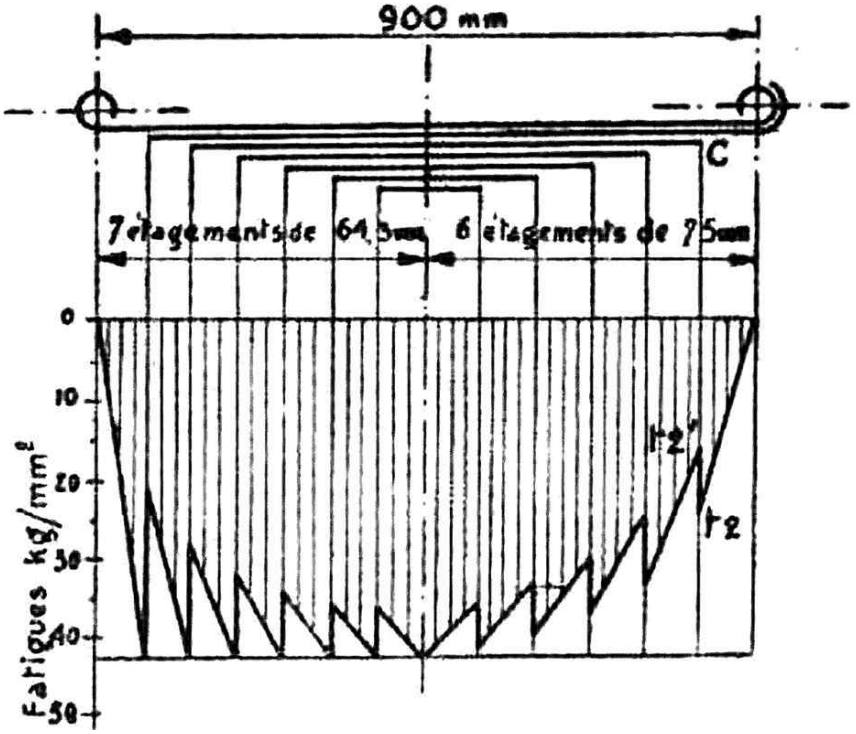


FIGURE 1.6. In the automotive engineering manual of Baudry de Saunier of 1900 the theory of a leaf spring with constant leaf thickness and length reduced by constant steps is explained. The structure is practically equivalent to a body of uniform resistance.

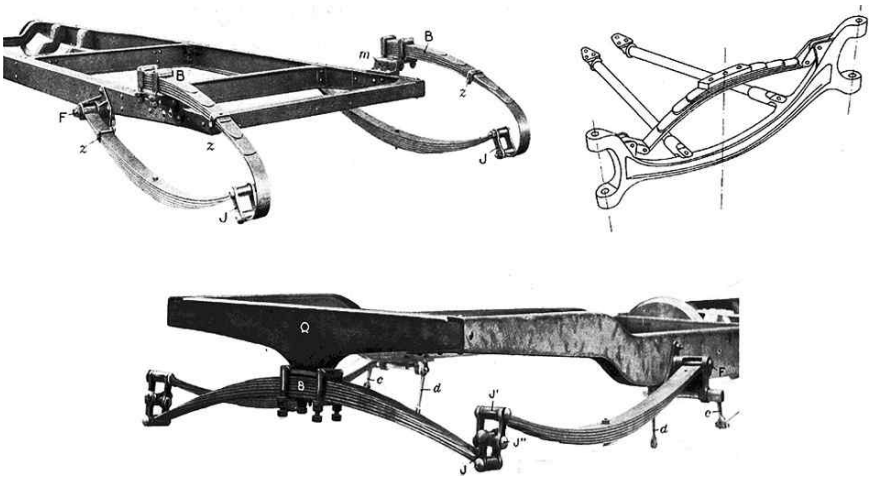


FIGURE 1.7. Some application schemes for leaf springs. On the top left is shown a $\frac{3}{4}$ ellipse solution; on the right the Ford scheme with a single spring for each axle; at the bottom a three leaf spring solution equivalent to that of the $\frac{3}{4}$ ellipse.

www.TechnicalBooksPDF.com

be used. Undesired motion of the axle in the longitudinal direction is avoided through the use of two dedicated thrust beams.

A convenient suspension should not only have a deformable elastic member, but also a damping element; otherwise a permanent oscillation of the sprung mass could take place as a result of an external force input. This cannot actually occur, because of internal friction; nevertheless numerous oscillations could be produced around the static position.

With leaf springs this problem was not addressed for some time, because of the high value of internal friction due to the relative motion of leaves; on the contrary, every effort was made to improve leaf lubrication in order to avoid annoying squeaks and permanent sticking because of oxidation. Nevertheless additional dampers were not initially considered. The first shock absorbers appeared in the 1910s, considered initially as accessories to be installed after the sale by demanding costumers interested in sport driving.

Many solutions were considered. Figure 1.8 shows the most interesting. The first four are friction shock absorbers; the first and the third work on the extension stroke only, while the second and the fourth work in the compression stroke as well.

Many engineers thought that damping forces were undesirable in the compression stroke because they increased the value of forces applied to the sprung

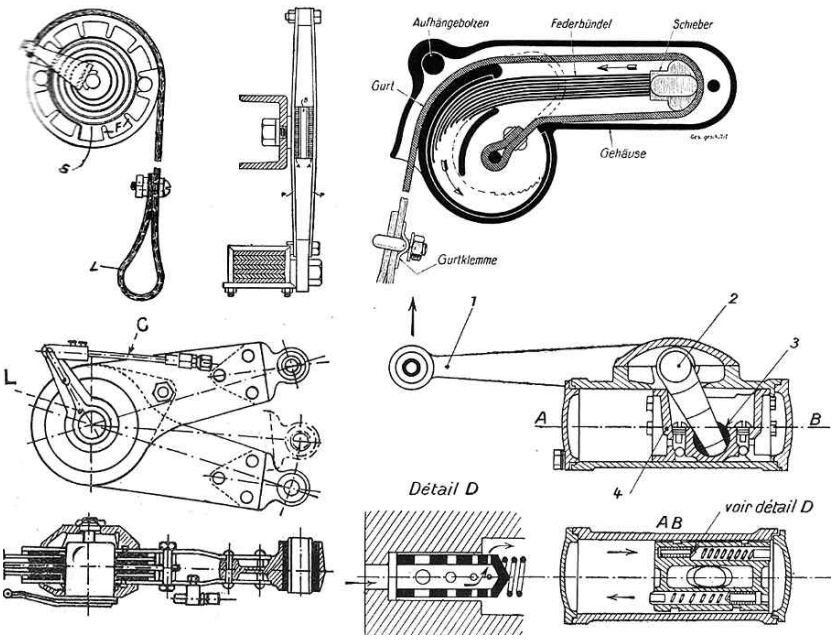


FIGURE 1.8. Some types of shock absorber, suitable for after market applications. The first three in the first row and the first in the second row work on mechanical friction; the last can also be adjusted. The fifth example refers to a hydraulic shock absorber.

www.TechnicalBooksPDF.com

mass; others thought that reducing the danger of oscillation was the most important priority. Only later was the importance of having a bilateral force proportional to suspension compression or extension speed understood and theorized.

A friction shock absorber has some properties that work against comfort. In fact, a dry friction force allows suspension motion only over a certain threshold value of force. It can happen that suspensions are completely blocked, by roads with little unevenness or by macadam pavement, in which case the only working elastic members are the tires. For this reason the fourth type of shock absorber features an adjustment screw suitable for changing the pressure force on the friction discs; this screw can be adjusted by the driver from the inside of the car, opening the possibility of reducing damping on good road or at low speed and increasing it under opposite conditions: The first example of controlled suspensions!

The last set of figures represents a hydraulic shock absorber, to be applied on the side beam of the chassis structure. This shock absorber works bilaterally with progressive forces depending on the suspension extension or compression speed. Modern telescopic shock absorbers implement this principle in a simpler way; these were introduced at the end of the 1920s.

We should not forget that the success of leaf springs was also bound to the reliability of the fabrication process for steel leaves. The microstructure of material could be improved easily by plane forging; a rupture of a leaf was never catastrophic, because leaves broke down one at a time and the consequent vehicle trim variation was easily detected. The necessary repair was simple and could be done by a competent smith. The arrival of coil springs was delayed by these facts; as a matter of fact leaf springs were initially applied to independent suspensions as well.

With the increase of vehicle speed some critical aspects of the rigid axle started to appear: The most important problems were weight, secondary deformations and *shimmy*.

The axle weight was remarkable; the amount of the unsprung mass as compared with the relatively high vertical flexibility of the tires could cause axle hopping on certain road, worrisome in turns and never suitable for comfort.

Secondary deformations were caused by braking (S-deformation) or by vehicle roll (different elongations of the two suspensions of the same axle). The first type of motion could cause annoying resonances of the entire driveline while starting up or braking on gravel or a dirty road.

The roll motion of the body, due to the centrifugal force in a turn, causes a different kind of longitudinal motion for the two points where the leaf springs are joined to the axle; in other words, the rigid axle steers, while the vehicle is turning. The architecture dictated by the transmission installation (fixed eye in front and swing shackle in the rear side) causes the vehicle to oversteer or to increase the path curvature because of the deformation; this fact could cause instability or spinning in curves that today appear undemanding.

www.TechnicalBooksPDF.com

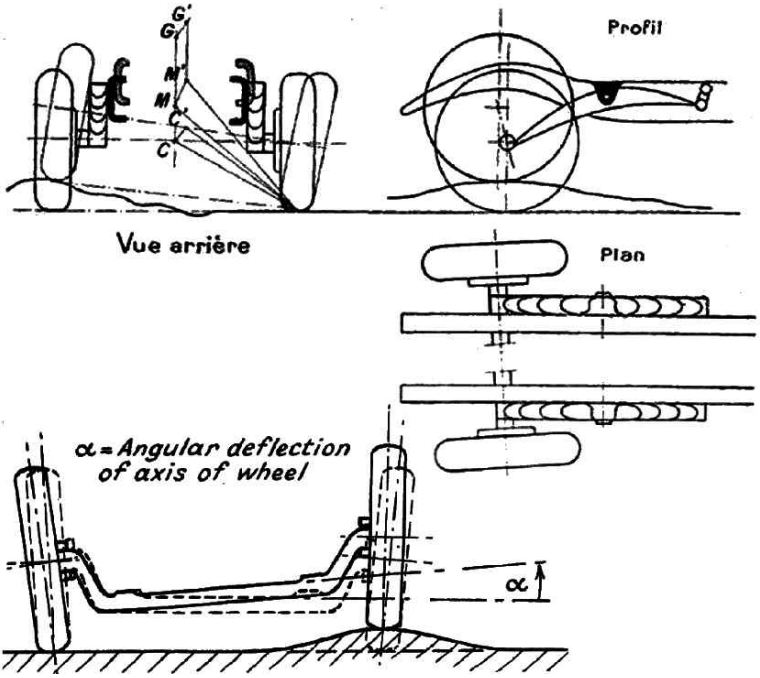


FIGURE 1.9. The upper scheme illustrates the axle steering of a Rolls-Royce suspension while negotiating asymmetric obstacles or in a turn. The lower scheme depicts the phenomenon of the *shimmy*.

The scheme at the top of Fig. 1.9 illustrates this behavior for a Rolls-Royce rear suspension; the advantage is that while driving on an asymmetric obstacle a self aligning steering action can be provided.

To avoid this inconvenience some rigid axle suspensions were built with two longitudinal linkages that modified the kinematic behavior of the axle; an example is given in Fig. 1.10, where two rods have the double function of avoiding the S deformation and of giving the axle an understeering behavior.

The lower part of the previous Fig. 1.9 explains the *shimmy* phenomenon typical of rigid steering axles; riding over an asymmetric obstacle imposes a certain roll speed on the axle. Because the wheels are rotating and had in old cars a significant inertia moment, a gyroscopic torque is applied to the steering mechanism; the elasticity of the steering wheel driveline (some elasticity was required to limit steering wheel vibration caused by the engine) could induce annoying oscillations. Because obstacles are primarily asymmetric this phenomenon was well known; it was solved in steering rigid axles by using additional shock absorbers, to be applied to the track rod.

A major advantage of the rigid axle was highlighted after the spread of independent suspensions; on rigid axles the body roll does not affect the camber angle of the wheel, because the wheel is always perpendicular to the surface of the road.

www.TechnicalBooksPDF.com

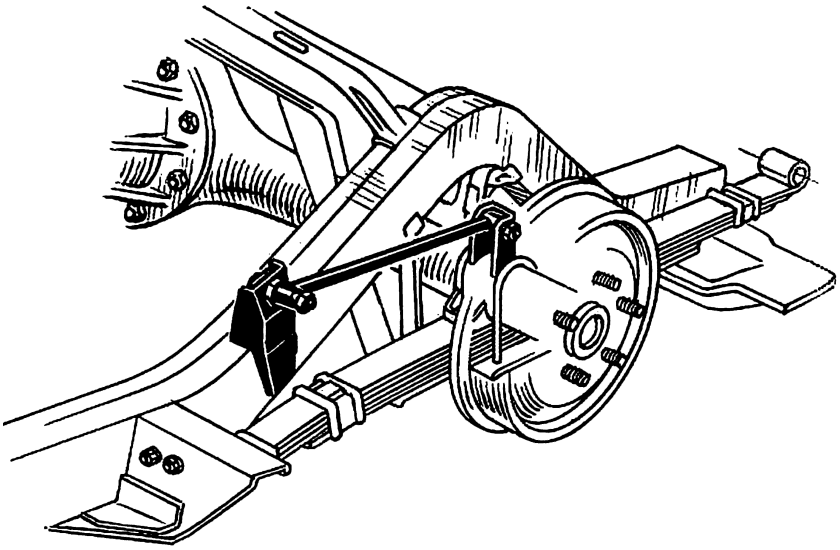


FIGURE 1.10. Example of a rigid axle with leaf springs, showing additional linkages fo reducing axle steering and S deformation.

1.3 THE INDEPENDENT SUSPENSION MECHANICAL LINKAGES

Rigid axle suspensions with leaf springs dominated the market for a long time and they began to be replaced by independent suspension only in the 1930s; nevertheless many examples of independent suspensions are older than this date, beginning with the very first, the already cited Mancelle.

Let us first consider steering suspensions.

Many examples of suspensions in which the wheel hubs are guided by vertical tubes can be found; this solution was perfected and produced by Lancia.

Probably the first application of this kind appeared in the Stephens of 1898, shown in Fig. 1.11, where the front wheels are guided by a telescopic steering fork inspired by a bicycle; they are sustained by a transversal leaf spring fit in the middle to the body.

This architecture was later considered by Sizaire & Naudin and Decauville, whom many authors quote as precursors to Lancia. In any case we doubt that these precursors had a clear view of the advantages they could claim.

It is interesting to note that many of the first advertisements for independent suspensions showed the wheel while driving over an asymmetric obstacle but never on a curved road. This suggests that the new architecture was developed to avoid the shimmy phenomenon while no attention was paid to the camber variation caused by roll angle; it also explains why independent suspensions were first applied to steering axles.

www.TechnicalBooksPDF.com

FIGURE 1.11. The vertical tubes suspension was probably introduced for the first time by Stephens in 1898.

The problem of camber angle variations was studied and solved only in the 1940s, when practical speed increased significantly.

Another input for change, we should not forget, was the advantages obtainable for the entire vehicle architecture. The front rigid axle, with its relevant lateral extension, could not be set too close to the ground; over the axle a certain clearance had to be provided to allow for the suspension compression stroke. As a consequence the engine had to be positioned at a relevant height over the ground; the front part of the body, the hood begun just behind the front axle and had a relevant vertical dimension.

The independent wheel suspension made this layout no longer necessary; the engine could be set down and forwards, between the suspension arms, with advantages for vehicle length and weight, reduction on the height of the center of gravity and consequent improvements in speed because the use of streamlined aerodynamic shapes.

Let us return to front independent suspensions with vertical tubes. Figure 1.12 shows a set of sketches drafted by Falchetto, the vehicle engineer who, under the guidance of Vincenzo Lancia, faced the problem of introducing an independent wheel suspension on the future Lambda.

The solutions considered on that occasion were almost all developed in the following years; the choice favoured the third in the third row from the top, probably because of the problem of shimmy, but there is no record of the decision-making process.

Figure 1.13 shows a picture of the final vehicle launched in 1922; the most important feature of this design, as we saw not completely new, was the solution

www.TechnicalBooksPDF.com

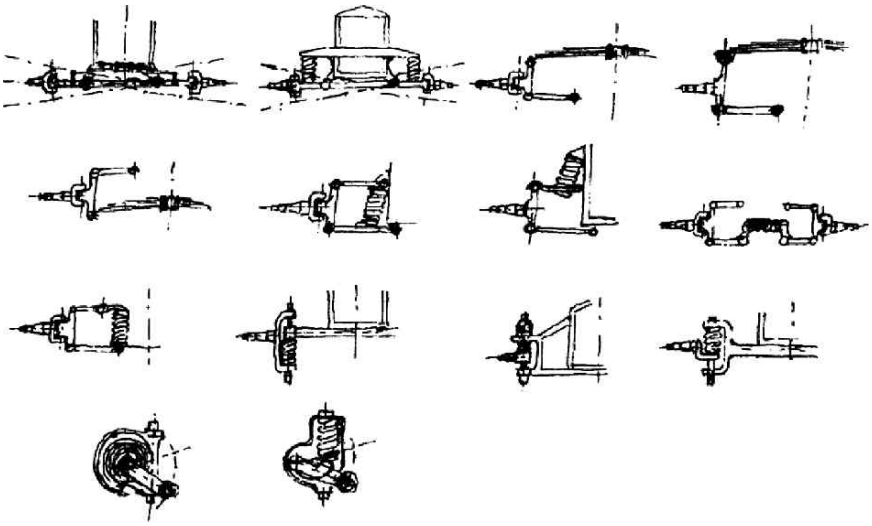


FIGURE 1.12. This table is evidence of the alternatives that were examined by Falchetto while designing the front suspensions of the Lancia Lambda. The architectures considered are almost all found on present cars.

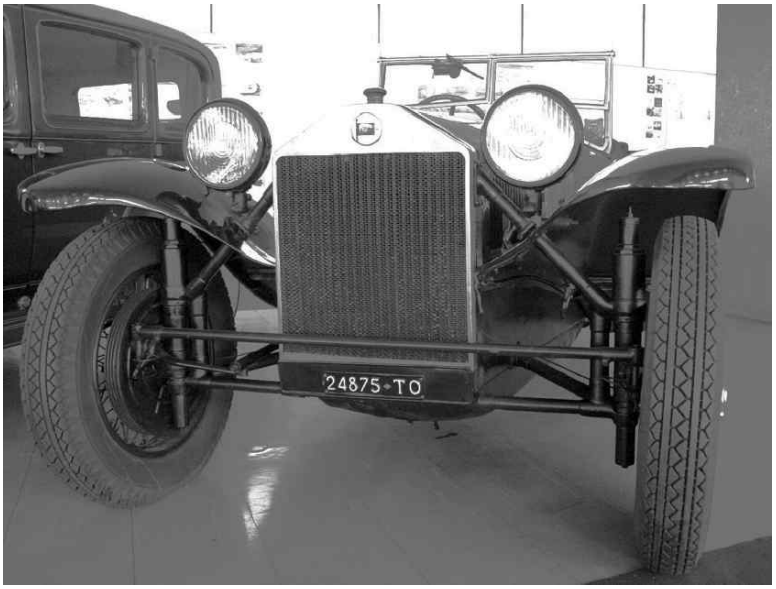


FIGURE 1.13. The front view of the Lancia Lambda shows the front suspension and the gate structure of connection to the body (Automobile Museum of Turin).

www.TechnicalBooksPDF.com

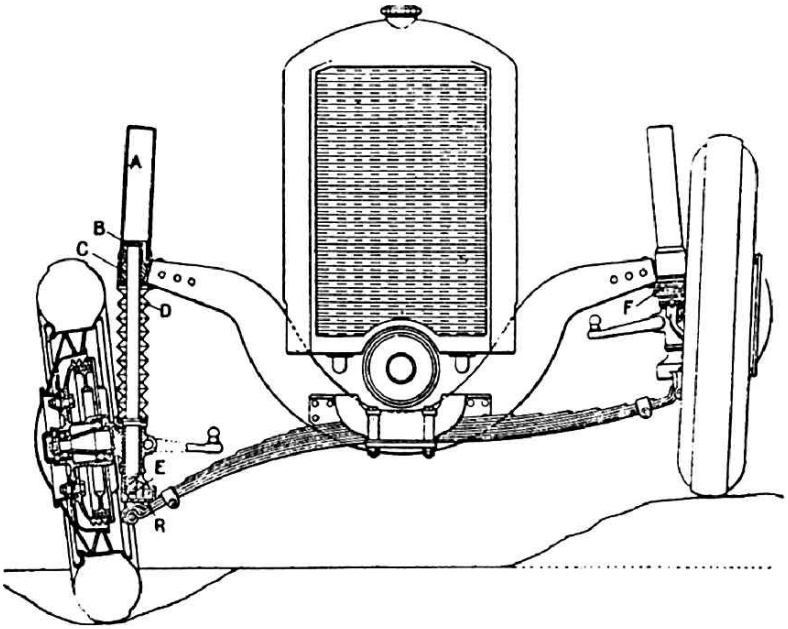


FIGURE 1.14. This front suspension was presented by Cottin-Deguttes in 1927. No other examples of this kind were developed.

to the problem of efficient lubrication of the sliding tubes by integrating, in a sealed element, guide, spring and hydraulic shock absorber.

The elastic element is now a coil spring; this is probably the first application in which the car weight is entirely sustained by a spring of this kind.

An interesting variation of this suspension family was presented by Cottin-Degouttes on a car shown in Fig. 1.14 and introduced in 1927; the vertical tubes suspension was modified, but we do not know the reason for this decision.

The tubes are slightly inclined with the top closer to the center of the car, while the elastic element is again a cross leaf spring, as in the Stephens. The leaf spring is articulated to the sliding strut; because the path of the tip of the spring is almost circular and the strut is coupled to the tube with spherical element, we can assume that the wheel can partially recover the camber angle variations caused by body roll.

Although the kinematic properties of this suspension can be similar to a double wishbone suspension, this kind of architecture had no industrial follow up.

A different scheme of front independent wheel suspension is attributed to Dubonnet, an important independent car designer, and was exploited by many car manufacturers, among them Fiat which applied this solution to production starting in 1935.

The suspension includes a sealed bearing element (cartridge) that integrates the helical spring in the same oil as the shock absorber, as shown in the detail on

www.TechnicalBooksPDF.com

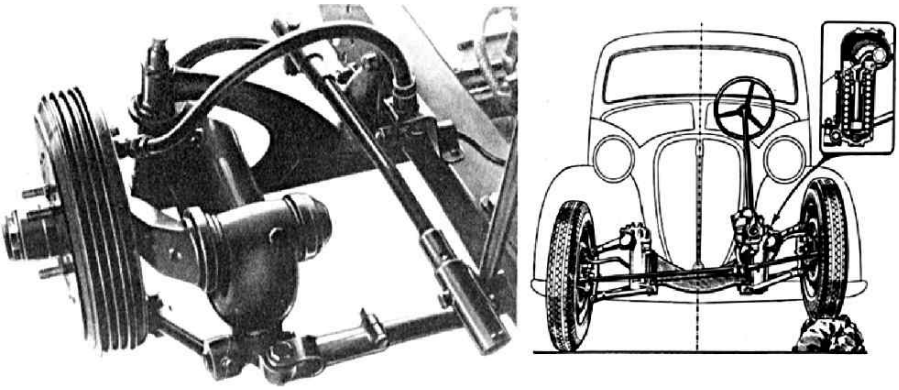


FIGURE 1.15. Two versions of independent front suspensions designed by Dubonnet and produced by Fiat, beginning in 1935. The suspension includes a bearing sealed cartridge, integrating the elastic element and the shock absorber.

the far right of Fig. 1.15; spring and damper work on a finger crossing through the cartridge. This finger is connected to one of the arms of a double wishbone mechanism.

The cartridge can be integrated into the suspension in two different ways. As we can see on the left of the same figure, the cartridge can be mounted on a non-suspended rigid axle through a king-pin; in this case the entire suspension is steering with the wheel and the steering mechanism is the same as for a rigid steering axle, making the motion of wheels completely independent. The kinematic behavior is almost the same as a vertical tube suspension, but the spring and shock absorber unit is easier to manufacture. The architecture is similar to that of a double trailing arms suspension.

The second alternative, shown on the right of the same figure, offers a cartridge flanged to a front cross member of the chassis structure. The suspension arms can swing but not steer; the wheel strut is mounted with knuckles to the oscillating arms. As a double wishbone suspension the steering mechanism is different and the track rod is replaced by an articulated system.

A suspension similar to the first scheme of double trailing arms is attributed to Porsche and is shown in Fig. 1.16; it was developed in 1931 and until the 1970s on the Volkswagen Beetle and other cars from this company.

The two articulated parallelograms are mounted at the end of a double tube structure that also build up the front cross member of the platform.

The upper arms are flanged to the same torsion bar, contained in the upper tube; this bar limits the roll angle. The lower arms are flanged to two different torsion bars, each of them flanged at the other end to the tubular structure; these act as elastic elements. The dimensions of this suspension are therefore constrained in the three directions.

While these suspensions were effective at reducing shimmy, they later demonstrated an undesired feature: They caused the wheel to have, during turns, a camber angle equal to the roll angle of the body.

www.TechnicalBooksPDF.com

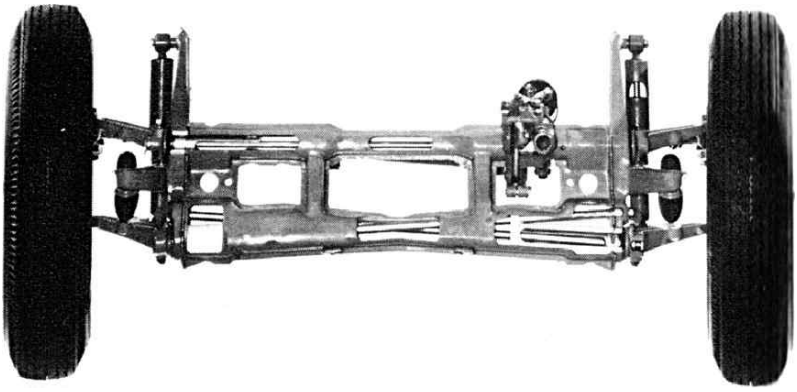


FIGURE 1.16. The Porsche suspension is made with two double trailing arms elements; it was developed in 1931 and produced by Volkwagen till the 1970s.

Bearing in mind that the roll angle, due to the centrifugal force, tilts the body to the outside of the curve, the wheels assume a camber angle so as to reduce the cornering stiffness of the tire. It would be better if the camber variation were opposite, but the ideal behavior is to have no variation under any condition.

To improve this undesirable situation double wishbone suspensions were developed with arms of unequal length; the shorter upper arm increases the camber angle when the suspension compresses and decreases when the suspension is extends. The value of the camber angle is such as to compensate partially for the effect of the roll angle.

One of the first examples of this concept was introduced by Studebaker in 1939 and followed by many car manufacturers until the present. A drawing of this suspension is shown in Fig. 1.17.

We would like to point out that the elastic element is integrated with the lower arm and is still made with a transversal leaf spring; as a matter of fact many manufacturers were not keen to abandon leaf springs. The reasons were many, chief among them reliability, but we can also see in this application that the leaf spring integrates the function of two elements: the arms and the spring. The reliability of the production process was likewise important and, last but not least, the existing investments for a high volume of production.

An interesting Fiat patent exist on this subject, applied to cars with rear engines, beginning in 1955; it fit the leaf spring to the body through two symmetric bearing points. With a suitable distance between the bearing points it is possible to obtain different elastic characteristics on symmetric and asymmetric suspension strokes; in this way, it is possible with a single spring element to obtain the function of the main elastic element and the anti-roll bar.

We can also observe in Fig. 1.17 that the Jeantaud mechanism present on previous steering systems has disappeared; the track rod can no longer be applied, because the distance between the articulation points of the track rod arms changes with the suspension stroke.

www.TechnicalBooksPDF.com

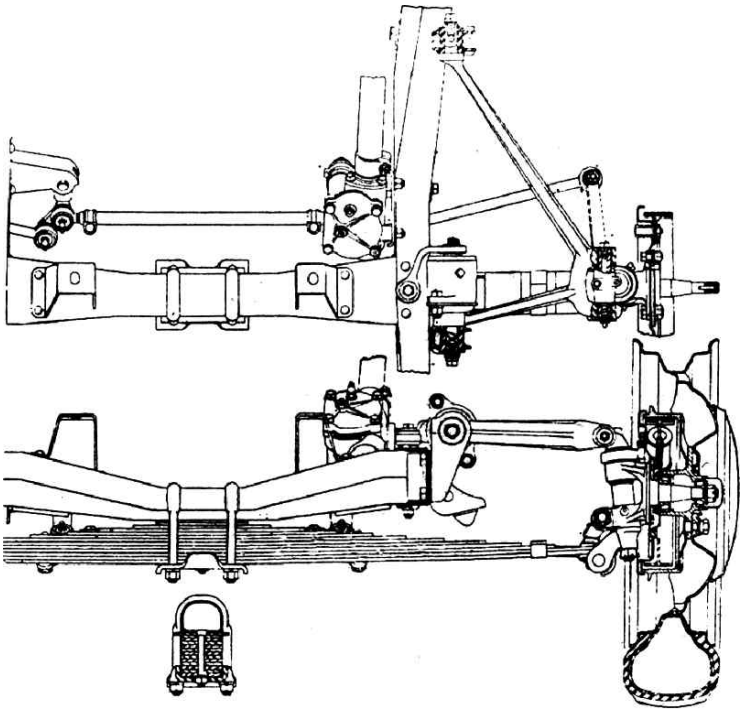


FIGURE 1.17. Front suspension made by Studebaker in 1939. The mechanism is the double wishbone type and features an upper arm of reduced length. The lower arm integrates the elastic function, being made with a leaf spring.

Before the advent of the rack steering system, applied beginning in 1960s, a small articulated parallelogram was employed whose rods were articulated to the body and rotated by the steering box; to suitable points on these rods were connected the two steering linkages.

This mechanism also approximates, with some error, the ideal Ackermann law; the presence of ever increasing sideslip angles, because of increasing speed makes this error less important.

The theoretical model of the behavior of tires under applied side forces and the derivative concept of sideslip angle were also developed during these years.

Double wishbone suspensions with different length arms spread quite rapidly in the following years and became almost universal in front axles during the 1960s.

Figure 1.18 represents a suspension with stamped steel low thickness steel arms and coil springs launched by Fiat in 1950 on the 1400 and applied with minor modifications on smaller models in the following years.

McPherson, a design engineer of Ford in the U.S., introduced the front suspension in 1947 that was named after him; it can be considered a double wishbone suspension with different length arms, where the length of the upper arm is infinite (see Fig. 1.19).

www.TechnicalBooksPDF.com

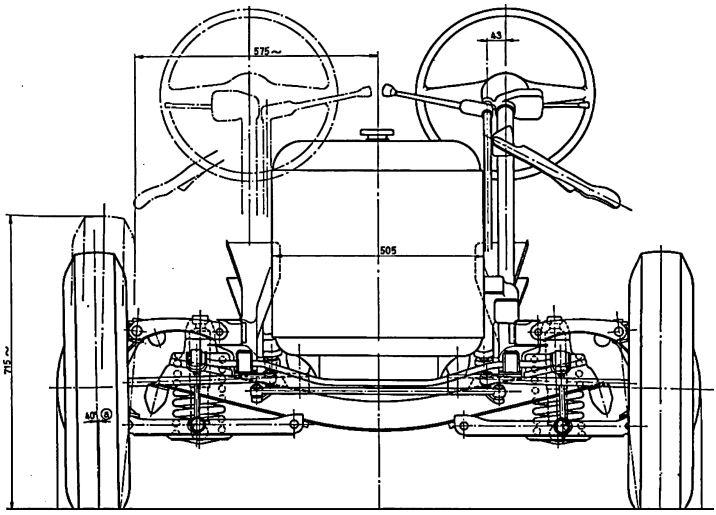


FIGURE 1.18. Double wishbone suspension adopted on Fiat cars in 1950; this solution was almost universal in the following years because of fairly good elastokinematic behavior.

This suspension should not be confused with a basic simplification of the double wishbone suspension for cost reduction; it later contributed to the diffusion of modern front wheel driven cars, because the lack of the upper arm left the necessary space for a transversal engine installation. This suspension spread rapidly beginning at the end of the 1960s, not only on front wheel driven cars built in those years but on numerous cars with longitudinal engines.

The second advantage of this suspension is that it simplifies the body structure thanks to the more rational distribution of connection points.

The reduction of kinematic performance as compared with the double wishbone solution is not very relevant and this solution is also applied to contemporary sport cars.

Honda in the 1980s introduced a final innovation in the front suspension architecture. It conceived a swan-necked wheel strut (see Fig. 1.20), which allowed the double wishbone suspension to also be installed on transversal front wheel driven cars. The descendants of the McPherson and Honda suspension today share the market, with the introduction of many improved details that we will omit for the sake of simplicity.

Rear suspension history should be much more complicated; the attempt to identify evolutionary trends risks oversimplifying the explanation.

It should be pointed out that the rigid axle, receiving an elastic member more sophisticated than leaf springs and additional linkages, had long life both on front and rear driven cars. On front wheel driven cars the weight of the axle was not relevant; in fact the simplified function, because of the absence of the differential and final drive allowed the use of tubular structures that were quite light.

www.TechnicalBooksPDF.com

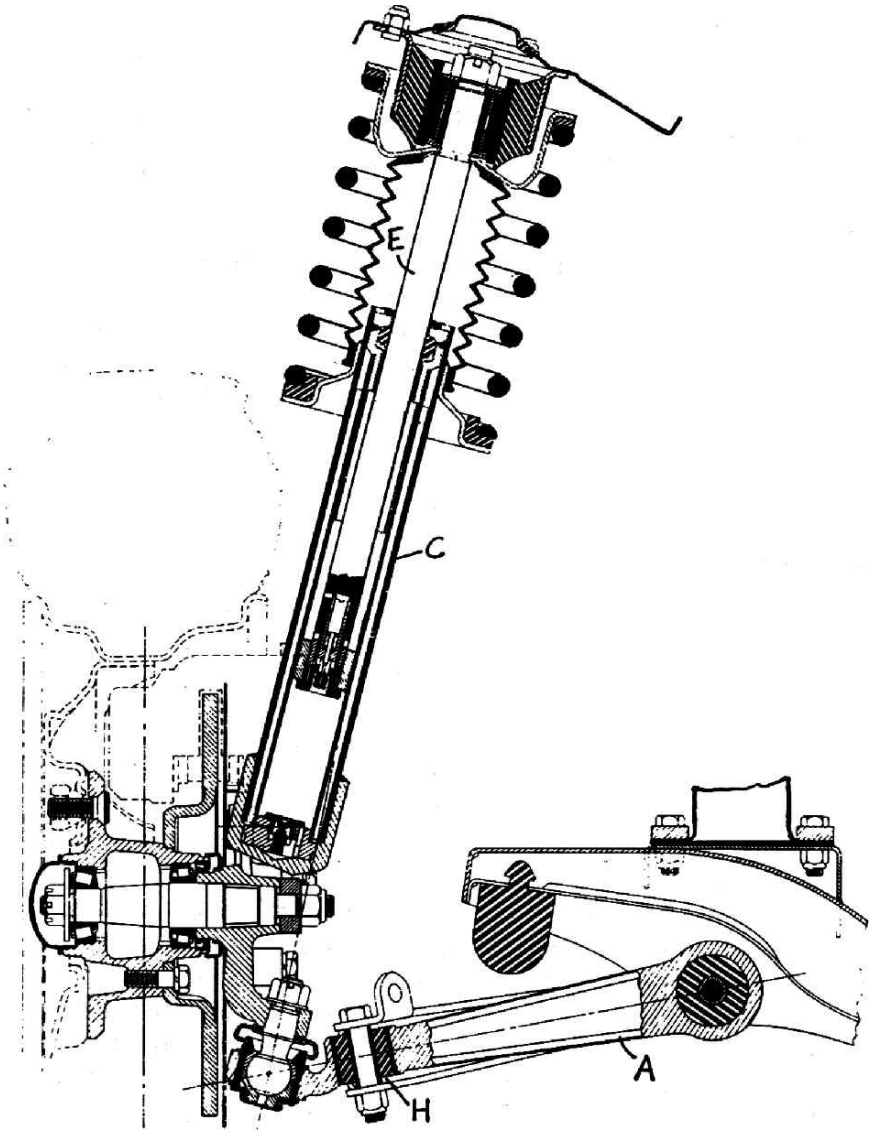


FIGURE 1.19. McPherson introduced on American Ford cars this kind of suspension beginning in 1947. The kinematic equivalent of this suspension is a double wishbone suspension where the upper arm has infinite length.

On rear wheel driven cars, particularly on luxury and sport cars adopting the rigid axle, the weight was reduced with suspended differentials, good kinematic behavior was obtained with coil spring and more complicated linkages.

A particularly elegant example of rigid rear axle is provided by Alfa Romeo. This scheme was adopted on different cars starting in the 1970s. It can be considered an improvement of the De Dion-Bouton suspension. Figure 1.21 shows this design.

www.TechnicalBooksPDF.com

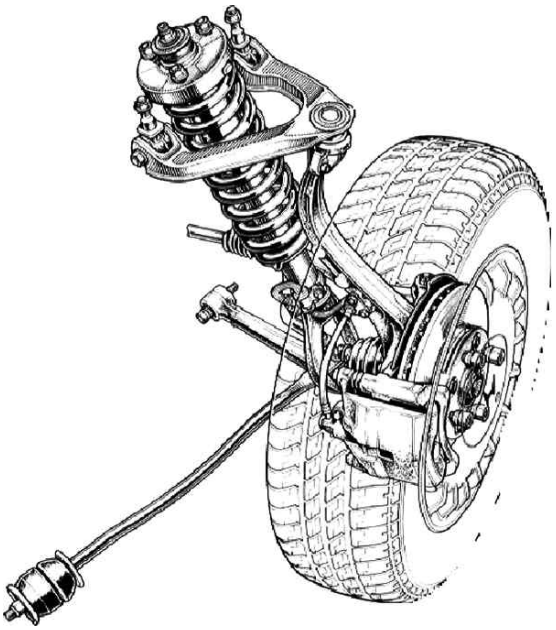


FIGURE 1.20. Honda front suspension with high double wishbone; high refers to the position of the upper arm compared to the conventional case.

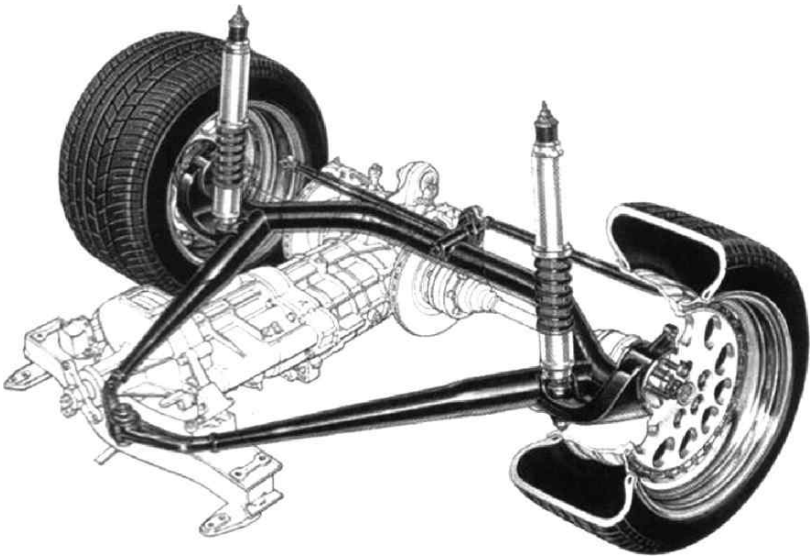


FIGURE 1.21. A particularly elegant example of a rigid rear axle is provided by Alfa Romeo. This scheme was adopted on different cars starting from the 1970s. It can be considered an improvement of the De Dion-Bouton suspension.

www.TechnicalBooksPDF.com

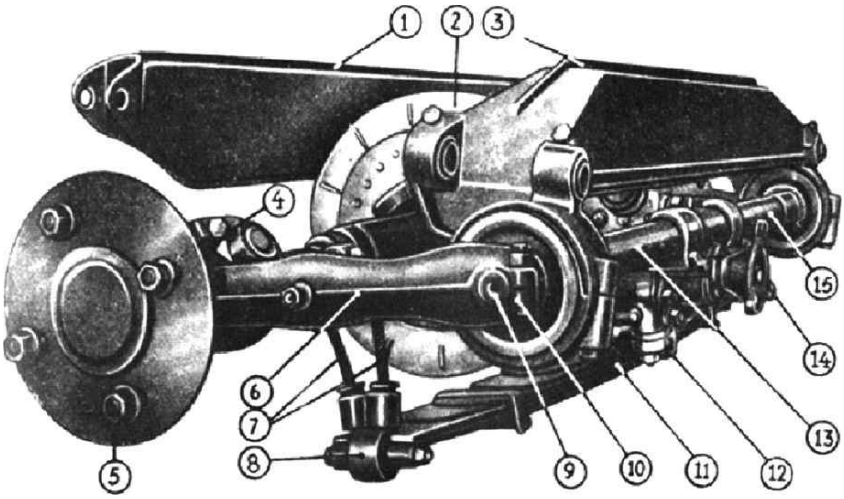


FIGURE 1.22. One of the first applications of an independent rear suspension with trailed arms appeared on the Lancia Aprilia in 1937.

A triangular structure building up the rigid axle is linked to a spherical joint in the front, which determines a precise position for the roll axis; suspension stroke and body roll do not affect axle steering thanks to the guidance given by a Watt mechanism in the back.

The need to obtain a well shaped and spacious trunk imposed the development of rear suspensions different from the rigid axle.

One of the first applications of an independent rear suspension with trailing arms appeared on the Lancia Aprilia in 1937. This suspension, presented in Fig. 1.22, is adapted for rear wheel drive and consists of two longitudinal arms with the articulation point in front of the wheel; the elastic element remains a transversal leaf spring with the addition of a transversal torsion bar.

Being an independent suspension, the differential and final drive is suspended on the body. This kind of suspension, in connection with the application of coil springs, had considerable diffusion up until the present, primarily on cars with front wheel drive.

Such suspensions enjoy the advantage of reduced space for linkages and elastic system; the disadvantage is the modest performance and the weight of the trailing arms.

Rear wheel driven cars with rear engine had a significant diffusion between the 1940s and 1960s; today they are no longer found, with the exception of some niche sport cars.

The reason for the diffusion of this architecture was the excellent interior roominess with contained exterior dimensions, thanks to the absence of the propeller shaft, at a time when economical and reliable components for the front wheel drive were not yet available.

www.TechnicalBooksPDF.com

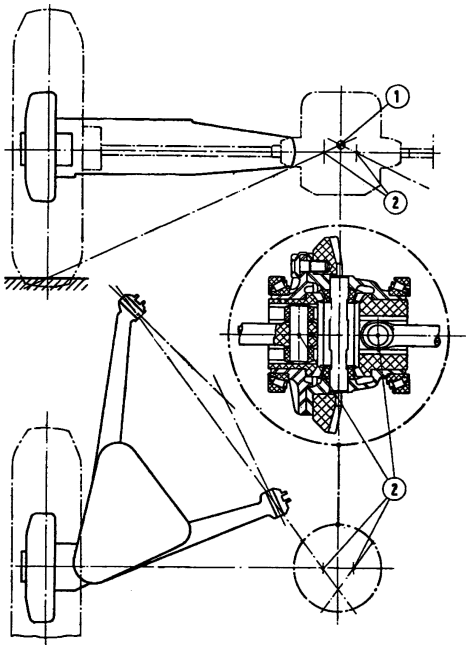


FIGURE 1.23. The rigid axle could not be applied because of compatibility with the rear power train. On the Fiat 600 of 1955 the semi-trailing arm suspension was chosen because it allowed the use of simple constant velocity joints.

The rigid axle could not be applied because of the short distance between power train and axle; so called semi-trailing arms were usually selected, because of their compatibility with inexpensive constant velocity joints.

In Fig. 1.23 is shown a scheme of the Fiat 600 of 1955. The traced dotted lines show how it was possible to have the suspension roll axis cross the differential; this condition is mandatory for avoiding the application of sliding constant velocity joints.

The camber angle relative to the ground does not change because of the roll of the body, but varies remarkably because of payload variations; the look of the unloaded car is characterized by a noticeable positive camber (wheel mean planes cross below the ground). In this condition rod holding is quite approximate. In addition, on bumpy roads the track changes continuously because of the suspension stroke, causing premature tire wear.

This kind of suspension was also adopted in many rear wheel driven cars with front engine because of its limited height; today it has been abandoned.

In 1969 Volkswagen conceived a new rear suspension quite suitable to front wheel driven cars: The so called semirigid axle or twist axle; in a short time it became one of the most widely diffused architectures.

www.TechnicalBooksPDF.com

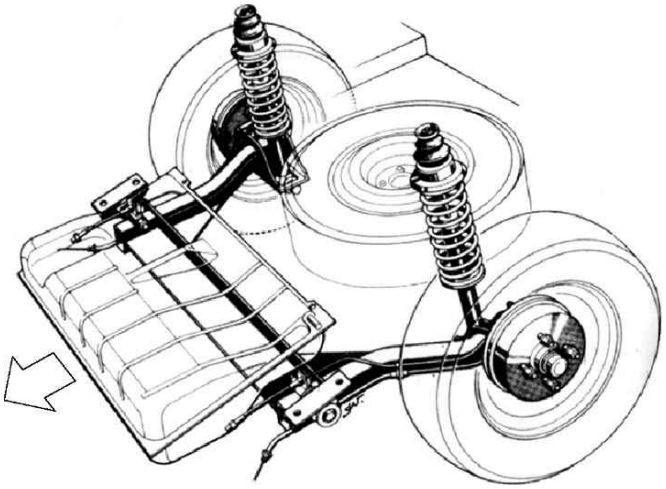


FIGURE 1.24. In 1969 Volkswagen conceived a new rear suspension quite suitable to front wheel driven cars: the so called semirigid axle or twist axle; in a short time it became one of the most diffused architectures.

Figure 1.24 shows the original application and demonstrates quite clearly the advantages produced on the car; the tank and spare wheel find their place between the arms and the cross member, leaving the trunk bottom low, flat and wide.

From this point of view alone, trailing arms are slightly better. They do not allow any camber angle variation relative to the body during suspension stroke; the steel sheets of the wheel wells can be closer to the tire profile and leave more space for the trunk at a given car configuration. Their elasto-kinematic behavior is poorer, in terms of the cornering stiffness of the tires, as we previously explained.

On twist axles the gate structure that bears the wheels is characterized by a cross beam made with a steel bar with open cross section, like a horizontal U, for example, open to the back. Using this feature the gate is flexible to torsion (differential suspension strokes), but stiff to flexion (side forces). The wheels are, therefore, effectively guided and do not change their camber angle respective to the ground very much during body roll.

We can conclude this historic overview of car suspensions by mentioning multilink architecture, which represents the pinnacle of suspension evolution. It was introduced during the 1960s on competition and sport cars, but was first mass produced by Mercedes in 1982; it is represented in Fig. 1.25.

If we consider the suspension solely from a geometric point of view, we can say that the suspension mechanism must leave the wheel but one degree of freedom, the stroke, not considering the wheel's rotation around its hub. If we

www.TechnicalBooksPDF.com

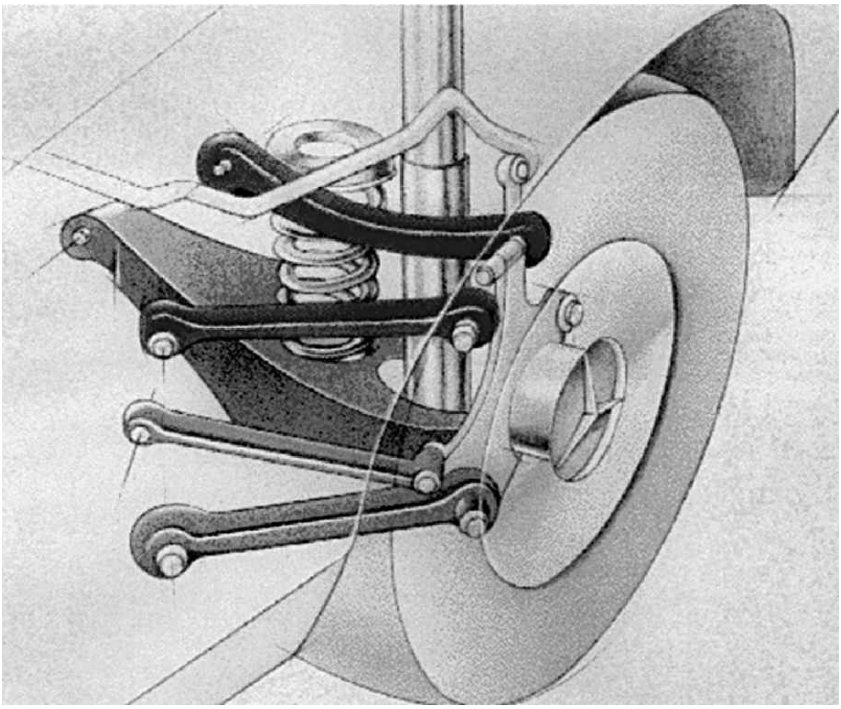


FIGURE 1.25. Multilink rear suspensions were first mass produced by Mercedes in 1982.

consider kinematic linkages with spherical joints only, we can apply a maximum of five linkages to reduce the six degrees of freedom of a free body in the space to one. The multilink suspension features five linkages to obtain the maximum number of adjustable parameters thus best approximating the ideal behavior.

With this arrangement we can minimize camber variation with reference to the ground. We can also induce angle variations capable of optimizing steering behavior and longitudinal flexibility. Optimum comfort can thus be obtained with no drawback on steering.

The multilink suspension family is very large; it is now found on almost all rear wheel driven cars with front engines; many members of this family are already present on large and medium front wheel drive cars.

The chapter on independent wheel suspension covers the last sixty years of automotive history, but evolution has not stopped. Although it is difficult to conceive new configurations of these mechanisms, many efforts are being made on the field of controls. This topic will be covered in the chapters dedicated to the state of the art.

www.TechnicalBooksPDF.com

1.4 WHEELS AND TIRES

Considering the important role wheels play on cars, we will extend our historical outline to much more ancient times.

For hundreds of thousands of years man lived without using any particular means of transportation. When he had to move an object, he simply lifted and carried it, if he was strong enough. If the object was too heavy, he arranged to drag it. It is likely that occasionally branches or other round objects were slipped under the load to reduce friction, but no evidence of this practice remains.

With the Neolithic revolution the need for transportation greatly increased at the same time that the practice of taming animals opened new perspectives. The development of agriculture created the need to transport seeds to the field and crops back to the homestead. The number of objects that were considered important and necessary for men to carry with them increased as a result of the new exigencies of village life.

It is a given that sleighs were used in northern Europe before 5000 B.C., and their use in other places at that time can be inferred. Sleighs and sledges can actually be used for transportation not only on snow and ice but also on grassland (American Indians used the *travois* well into the XIX century), deserts and sometimes even on rock.

It is impossible to state when a sledge was mounted for the first time on a pair of wheels or who instigated this technical revolution. Ancient wheels were made primarily of wood, so that little direct archeological evidence would remain.

In about 3500 B.C. the potter's wheel was introduced in order to produce pots with axial symmetry. The use of the potter's wheel can be inferred from the marks left on pots made with it. The supporting wheel for vehicles is thought to have originated at about the same time.

The most ancient evidence of a wheeled vehicle is a pictogram on a tablet from the Inanna temple in Erech, Mesopotamia. This document dates back to slightly later than 3500 B.C., and includes a small sketch of a cart with four wheels, together with a sketch of a sledge (Fig. 1.26a).

The vehicle shown in Fig. 1.26b has two features typical of all vehicles for more than a thousand years: The wheels are discs made from three planks of wood, with the animals harnessed to a central shaft. This uniformity of wheel types and driving systems, particularly when compared with the great variety of vehicle structures, has led to the opinion that the wheel was *invented*, or better, developed, in a specific place and then started a slow diffusion throughout the ancient world. In various places where the new vehicle was introduced, the local type of sleigh was adapted to it using the standard wheels and harness.

Where the wheel was first developed is not known, but it can be inferred that it was in Southern Mesopotamia, where the wheel was definitely used about 3500 B.C. The spread of the wheel was quite slow. Evidence of its use dates from 3000 B.C. in Elam and Assyria, 2500 B.C. in Central Asia and the Indus Valley, 2250 B.C. in northern Mesopotamia, 2000 B.C. in southern Russia and Crete,

www.TechnicalBooksPDF.com

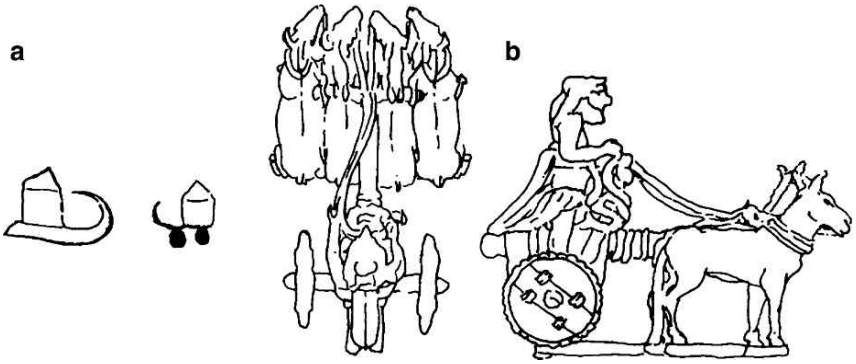


FIGURE 1.26. (**a**) Pictogram on a tablet from the Inanna temple in Erech, Mesopotamia. The document dates back to slightly later than 3500 B.C., and includes a small sketch of a cart with four wheels, together with that of a sledge. (**b**) Copper model of a war cart, driven by four onagers, found in the tomb of Tell Agrab, from the third millennium B.C.

1800 B.C. in Anatolia, 1600 B.C. in Egypt and Palestine, 1500 B.C. in Greece and Georgia, 1300 B.C. in China and about 1000 B.C. in northern Italy. Some centuries later it reached northern Europe.

It is impossible to know from ancient pictures whether the axle turned along with the wheels or was stationary. The fact that the central hole of the wheel disc was round has little meaning as a circular hole can also be explained by the ease of construction. It is likely that both solutions were used, as there are still people who use these primitive technologies today.

It is, however, likely that the wheel did not derive from the roller: The types of wheels used would rule that out, and it is likely that, in the mind of the ancient wheel maker, the wheel and the roller had little in common.

The need to build lighter wheels for war chariots probably led to the development of the spoked wheel, which is much more efficient for such use. The wheel with spokes was probably first used about 2000 B.C. and by 1600 B.C. had reached its fully developed form, particularly in Egypt. The central part of a wheel of that type (with eight spokes) is shown in Fig. 1.27a. It is part of a chariot dating back to 1350 B.C., found in a tomb near Thebes. The spokes are fitted into the hub; the felloe is usually built in various parts but some examples of felloes in one piece, bent into a circular shape, were found as well.

A wheel which seems to be a stage in the evolution between the disc and the spoked wheel is shown in Fig. 1.27b. The provenience of this wheel leads one to believe that it was simply an attempt to copy a spoked wheel by a wheelwright familiar with disc wheels; but wheels of the same type are represented on more ancient Greek paintings.

In many cases, even in very ancient times, wheels used a hoop or tire or at least some device to strengthen the rim. Some disc wheels have a wooden rim in one or more pieces. Sometimes the rim of the wheel is inlaid with copper

www.TechnicalBooksPDF.com

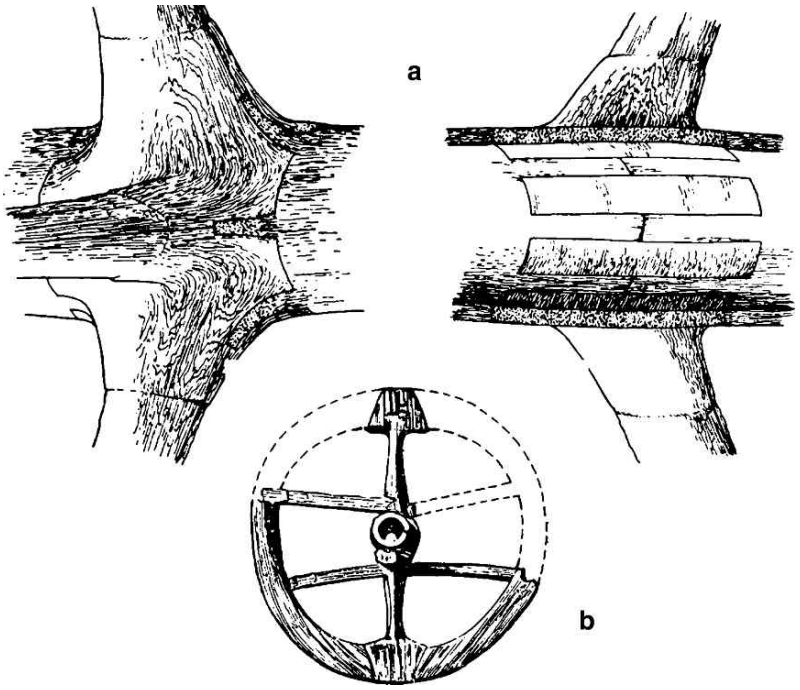


FIGURE 1.27. (**a**) Cross section and view of the center part of a wheel of an Egyptian cart from 1350 B.C. found in a tomb near Tebe. (**b**) Wheel found in Mercurago, on the River Po flat, probably dating to about 1000 B.C.

nails, to reduce wear or perhaps to hold a leather tire in position. Certainly many Egyptian war chariots had wheels covered with leather. In some pictures, even very ancient ones, something that looks like a metal tire can be seen. The evidence for such practice is however much more recent, dating back to about 1000 B.C. These metal tires were built in various parts, welded together and then shrink-fit to the wheel.

As previously noted, only after animals were tamed could wheeled vehicles be propelled in a proper way. In Mesopotamia both transportation vehicles and war chariots were pulled by onagers. Oxen were doubtless used for transportation as well.

The spoked wheel, which appeared about 2000 B.C., accompanied the use of horses to drive war chariots. It is not known where horses were first tamed and used for that purpose, but the scarce archeological evidence indicates that it likely happened in north-east Persia, and that from that region the use of horses spread throughout the ancient world, from China to Egypt and Europe.

The structure of an Egyptian chariot of the XV century B.C. is shown in Fig. 1.28. It represents without a doubt the best state of the art of its times, one that remained unchanged for centuries.

www.TechnicalBooksPDF.com

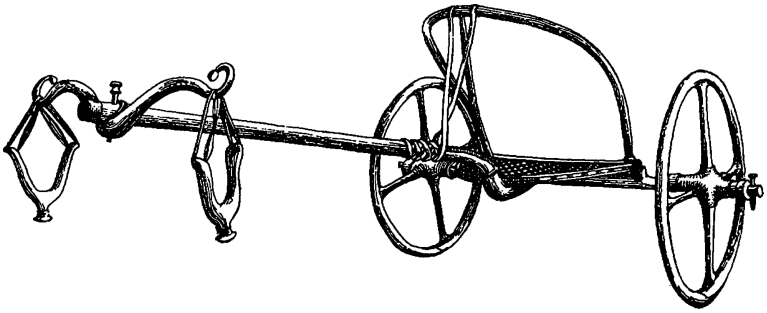


FIGURE 1.28. Egyptian war chariot found in a tomb near Tebe (XV century B.C.).

The progress from the Sumerian vehicle shown in Fig. 1.26b is great, and if the greater power of the two horses compared with that of the onagers is considered, it is easy to understand why some historians ascribed to the use of this weapon the expansion of the Hittites in Anatolia, the Achaei in Greece and the Hyksos, who in the XVIII century B.C. invaded Egypt, teaching the new technology to the Egyptians.

Chariots became obsolete when the knowledge of riding became wide-spread. Donkeys were used as pack animals and for human transportation in the third millennium B.C. and horses were surely used in the same way.

In Europe only the Celtic tribes continued to use war chariots, which were first carried north of the Alps by the Etruscans. Celtic wheelwrights learned the art of building wheeled vehicles and made significant progress.

The remains of the wagon found at Dejbjerg are shown in Fig. 1.29a. It is the first example of a wagon with steering on the front axle, but it can be considered an articulated vehicle made by two chariots. It is, however, unlikely that this solution was actually used for transportation; it looks more an insulated example for ceremonial (burial) purposes. At any rate it incorporated other interesting features, such as the wooden roller bearings in the hubs.

The wheel shows construction details we will find on more modern vehicles: Particularly interesting are the roller bearings, completely made of wood, in wheel hubs, as shown in detail b.

So-called artillery wheels were commonly applied to carriages and coaches at the end of XIX century and on many of the first cars; the hub of this wheel is shown in the upper part of Fig. 1.30.

The wheel is made by wooden spokes, whose base is shaped like a sector of the hub. They are clamped by two metal flanges which also build up the outer ring of the bearing; the rim is made of wooden arches pressure fitted in a steel ring which forces the spokes into the hub. Spokes are perpendicular to the hub with a certain inclination according to a bevel structure.

This kind of design increases the lateral robustness of the wheel and gives it the necessary radial flexibility to allow for hot pressure fitting of the rim.

www.TechnicalBooksPDF.com

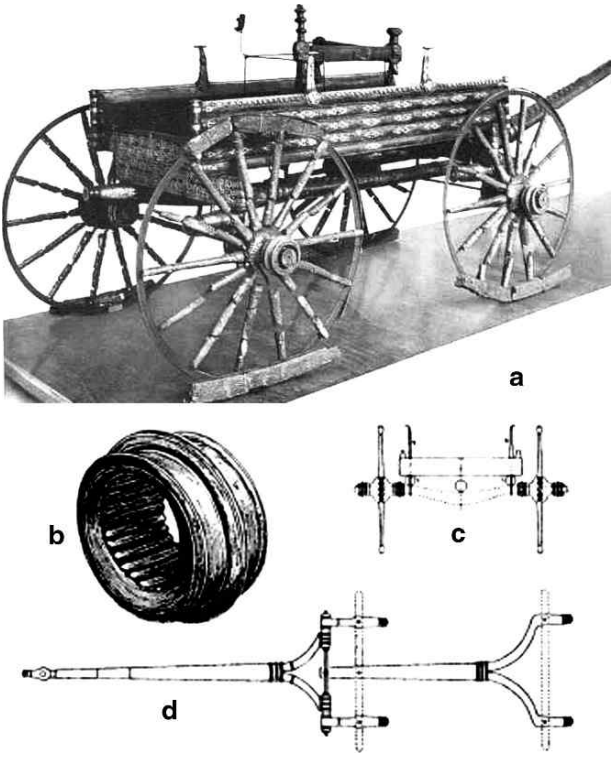


FIGURE 1.29. Celtic carriage from the I century B.C. found near Dejbjerg: (**a**) picture, (**b**) wooden roller bearings in the wheel hub, (**c**) cross section, (d) view.

This shape required a suitable inclination of the wheel rotation axis so as to have the spokes working correctly under the vehicle weight; this angle (lower sketch) was called the camber angle (*carrossage* in french).

On metallic wheels this kind of inclination was no longer necessary, but the name remained.

On this kind of wheel a solid rubber rim could be mounted. Other applications showed a cycle type wheel with solid rubber tire, like the cars of Daimler and Benz, shown in Fig. 1.4.

In Fig. 1.31 many of the fundamental evolutionary stages of the pneumatic wheel are shown; we will speak of the tire itself later.

The first wheel on the upper row, in use until the 1910s, was similar to the artillery wheel just discussed; instead of a metallic rolling rim, it offered a channel ring able to receive the pneumatic tire. These wheels cannot be easily dismounted from the hub; initially, a punctured tire could be changed only directly on the lifted car, with enormous difficulty.

The separable channel rim, shown in the second figure, makes changing the tire easier, allowing one or more spare rims with inflated tire on board.

www.TechnicalBooksPDF.com

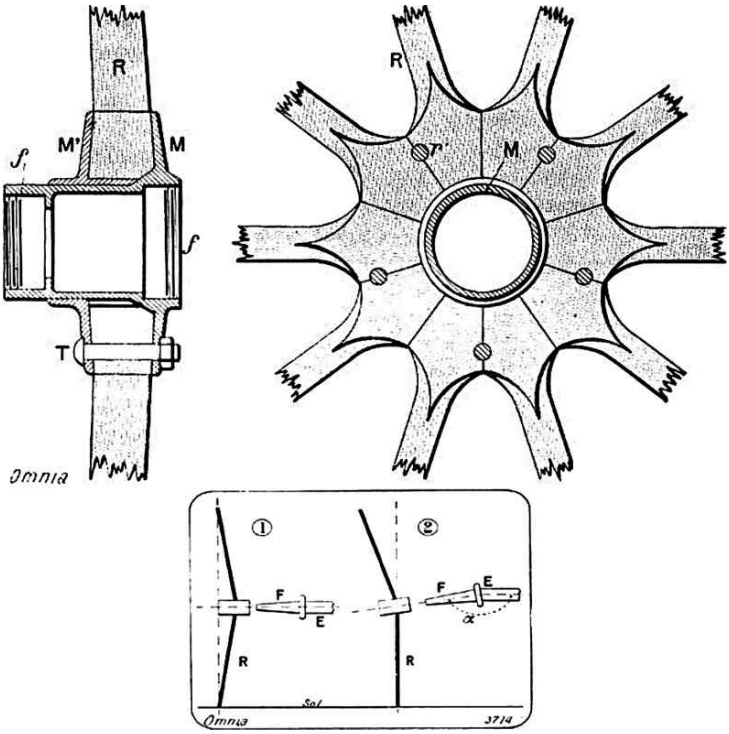


FIGURE 1.30. So-called artillery wheels were commonly applied on carriages and on the first cars. Wheel spokes are perpendicular to the wheel axis but are slightly inclined, following a bell shape. This shape required the wheel to be mounted at a certain inclination, with respect to the vehicle, so that the spokes work correctly. This angle was called camber angle (*carrossage* in french).

Solid metal wheels first appeared in the 1920s; the mechanical resistance was impressively increased. On this occasion the wheel was modified in order to be separable from the hub through a bolted flange, the so called Sankey wheel. Changing a punctured wheel became much easier.

Cycle type spoke wheels were used until the 1950s because of their low weight, especially on sport and luxury cars; the first type shown presents a conventional flange hub, while the second features a rapid dismounting Rouge Whitworth joint. This is characterized by a spline coupling, between wheel and hub, held in place by a large butterfly nut; this feature became a stylistic sign of fine cars.

In comparison with the first car of Daimler and Benz, we can observe on more modern cars, as in Fig. 1.32, a different spoke arrangement, now in two rows, with a different geometric configuration.

The first row, inside with reference to the car, has spokes with radial inclination only; these spokes transmit braking and driving torque and their inclination is motivated by the fact that the spoke can transmit forces only along its axis.

www.TechnicalBooksPDF.com

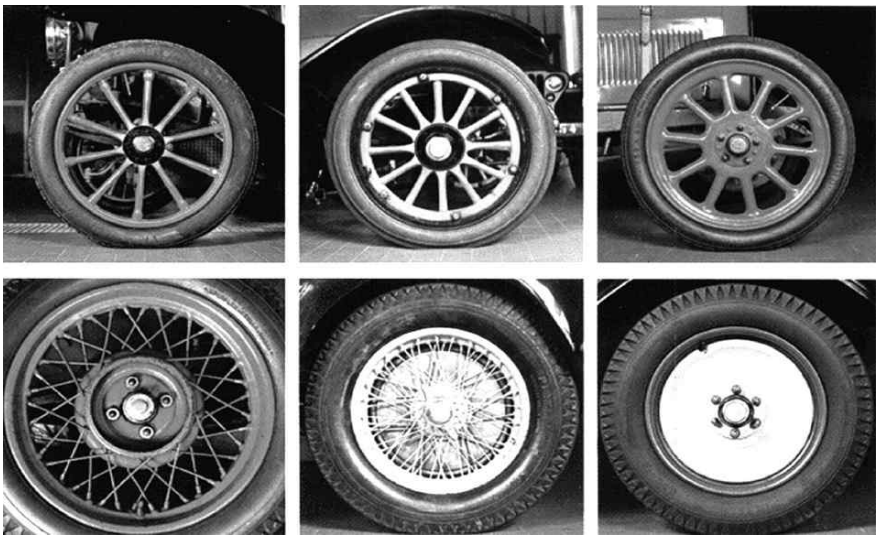


FIGURE 1.31. Evolution of the car wheel. The first two in the upper row are made of wood with metallic reinforcements; the second one has dismountable rim. The third is the Sankey wheel, made entirely of steel and separable from the hub, a solution also found on modern wheels. On the second row two solutions are shown with cycle type spokes; the second has a rapid dismounting fit. The last image shows the Michelin disc wheel scheme, used also on present cars.

The second, outside row provides for spokes on a bevel surface. Their orientation is suitable, along with those of the first row, to transmit lateral forces applied to the wheel.

Returning to Fig. 1.31 we see, as a last example, the Michelin wheel, which includes a stamped steel disc welded to the channel rim; this solution, after a series of improvements, is still in use.

If everyday life without a car is inconceivable, even more so is a car without tires.

The tire is not as old as the wheel, but is certainly older as the automobile.

Thomson is remembered as the inventor of tires, which were described in an English patent of 1845; the invention was designed to improve the ride comfort of coaches and to reduce their rolling resistance: The initial proposal provided for a leather sewed lining and a cloth tube treated with rubber.

Although not put into practice, this idea was proposed again by Dunlop, who filed a similar patent in 1888, without knowledge of Thomson's priority; this time bicycles could profit from this invention, which began to be applied extensively. As happened in the legal dispute of Otto regarding the internal combustion engine, Dunlop's patents were invalidated in 1890, because of Thomson's priority. This fact did not affect Dunlop's business negatively, but was beneficial to those who wanted to enter the same market contributing to a rapid development of this technology.

www.TechnicalBooksPDF.com

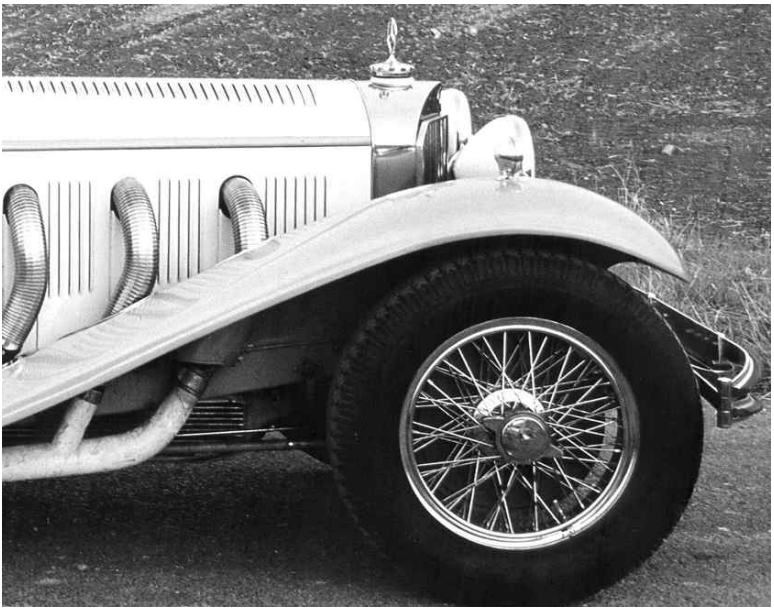


FIGURE 1.32. A spoke wheel with Rouge Whithworth nut; spokes are arranged in two rows. The inside row has spokes with only radial inclination; these spokes transmit braking and driving torque. The outside row provides for spokes on a bevel surface to carry lateral forces.

Only in 1980 were semi-tubular tires introduced; these could be adapted to a channel rim with a steel cord ring and allowed easy dismounting.

Michelin Brothers started to supply dismountable tires on cars, beginning with the Paris-Bordeaux race of 1895.

This could be thought of as the birth of the tires on automobiles, one that received many technological contributions by various manufactures.

A key invention for the spread of tires was rubber vulcanization, after an 1844 patent of Goodyear . Before the invention of this chemical treatment, probably discovered by chance, rubber could not maintain a stable shape and wore-out quickly depending on temperature.

Many further improvements allowed tires to reach their present level of performance; we recall the application of synthetic fibers to tire cloths, initially of line and cotton; these were made of rayon in the 1930s, later of polyesters and of kevlar from the 1970s on. A second major evolution refers to the radial texture of cloths, formerly diagonal, beginning in 1960s.

Other improvements involve the chemistry of rubber and the introduction of synthetic elastomers; on this subject we will not supply further details.

A final evolution regards the geometric dimension of tires, characterized by the channel diameter D, by the tire section width W and by the so called *aspect ratio*, which we will explain later, as the ratio between the radial dimension of

www.TechnicalBooksPDF.com

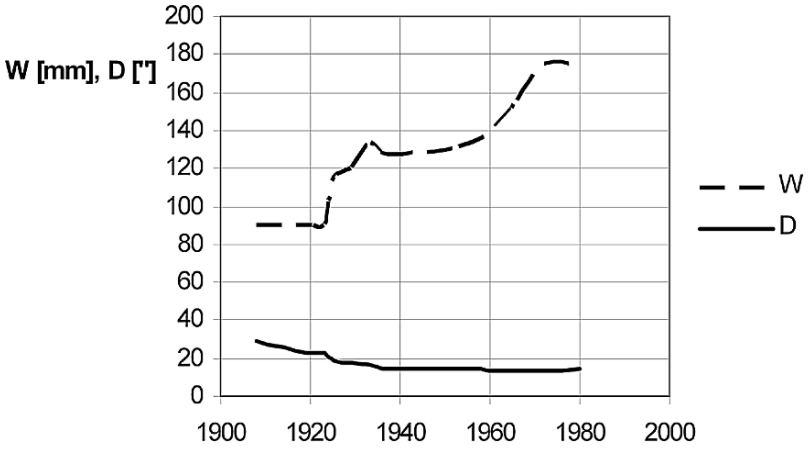


FIGURE 1.33. Evolutionary trend of tire dimensions on a car of about 1,000 kg of mass; the evolution is shown by the rim diameter D ["] and the section dimension W [mm].

the tire and its section width. The aspect ratio wasn't standardized, but was set around 100%; later, to improve comfort, tires became wider and the need to reduce car height and weight reduced the tire diameter. Aspect ratios began to be reduced.

Figure 1.33 shows the evolution of these dimensions for a series of Fiat cars, with curb mass in the range of 1,000 kg; the rim diameter decreases from 25" in 1910 to 13" in 1970, increasing slightly in the following years; the section width doubles from an initial value of 90 mm to about 180 mm on modern cars.

The aspect ratio, not shown in this diagram, had a step by step evolution; an initial value of about 100% became 80% in the 1940s, to descend again to 50% in the 1970s: In these years tires conserved their outside diameter, while the rim diameter gradually increased.

On more recent cars the outside diameter tends to increase.

We should finally point out that since the beginning of automotive history many patents on *elastic wheels* have been filed, where elasticity was obtained by mechanical elastic means rather than by the elasticity of the tire; these are often called *non-pneumatic tires*. Diffusion of tires has made this kind of wheel disappear for about a century, except in the wheel of the LRV of the Apollo Project (see Appendix C). Quite recently Michelin has presented an elastic non-pneumatic wheel called *tweel* (contraction of tire and wheel) that could lead to a real innovation in the future (Fig. 1.34).

1.5 BRAKES

The evolution of brakes is tied to increase in speeds and traffic density, both of which impose the need to stop the car in a short space; it is reasonable to think that braking the first cars in the traffic of that time was not a serious concern.

www.TechnicalBooksPDF.com

FIGURE 1.34. *Tweel*: an elastic non-pneumatic wheel recently developed by Michelin.

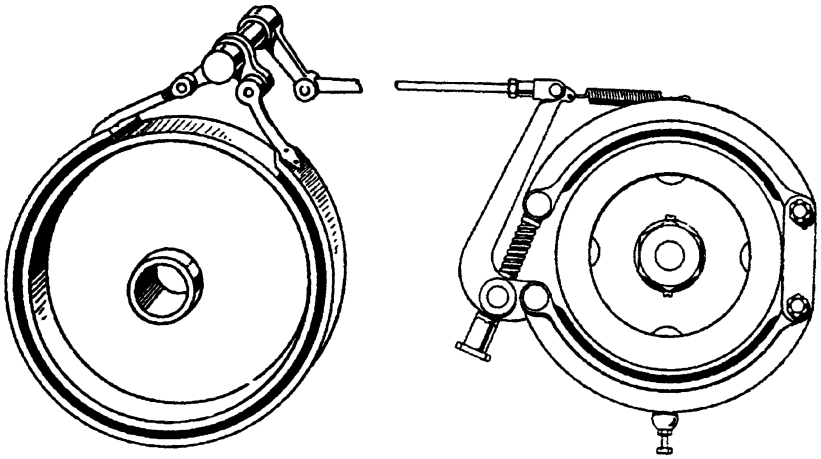


FIGURE 1.35. Band brakes and external shoe brakes, to be applied on the transmission line before the differential and final drive.

The earliest cars feature brakes that work on the external surface of the wheel with moving shoes, like those on horse drawn carriages and railway cars; this simple device was not suitable for tired wheels because of the potential wear.

For this reason external shoe brakes and band brakes, like those shown in Fig. 1.35 were introduced; they worked on the transmission line only. The choice was motivated by the need to exploit the differential for uniform distribution of the braking force, difficult to achieve with mechanical linkage control. Such brakes were also virtually insensitive to machining tolerances.

Brake liners were missing, but some examples of *Ferodo* liners, a synthetic material developed by Frood were already present on shoe brakes in coaches; this practice spread on a large scale in the 1920s.

Drum brakes were developed later and were applied on rear wheels only; they were more difficult to machine but showed a superior capability of dissipat-

www.TechnicalBooksPDF.com

ing heat generated over long braking distances. Band and shoe brakes were not initially abandoned, but continued to be applied on the transmission as supplementary or parking brakes.

Disc brakes were introduced into series production by Jaguar in 1953.

Drum brakes were controlled by pedal and were usually designed to stop the vehicle; the transmission brake, on the other hand, was controlled by a hand lever which could be stopped in the desired position. It was usually designed for continuous braking on downhill slopes and for use as a parking brake.

It should be noted that the position and function of driver controls was informally standardized only in the 1930s and the solution initially adopted could be much different; one system that received a certain diffusion which is seemingly unjustified today, provided for combined control of brake and clutch to avoid stalling the engine when braking ceased.

The application of brakes to the front axle began only in 1910, thanks to a luxury car manufacturer: Isotta Fraschini.

This innovation did not regard the brakes per se, which were similar to other rear brakes, nor the discovery of their usefulness, but the control system for steering wheels, as shown in Fig. 1.36.

Brake controls were in fact mechanical, as already shown in Fig. 1.35. They were made by linkages and cables put in tension by a pedal; in drum brakes the internal shoes were expanded by a cam. What had to be avoided, at any rate, was that wheel steering would affect braking; the patent of Isotta Fraschini

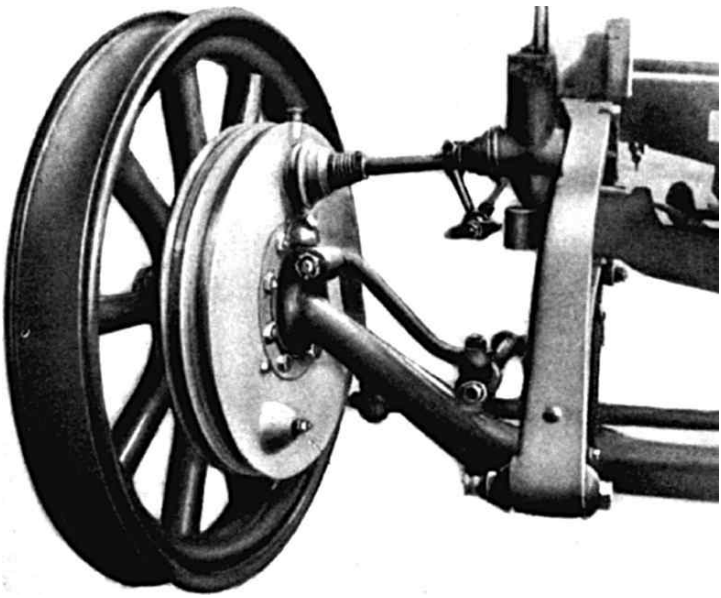


FIGURE 1.36. Completely mechanical control of front wheel brakes; the contact point of the cam, actuating the expansion of shoes, is near the king pin axis of the wheel.

www.TechnicalBooksPDF.com

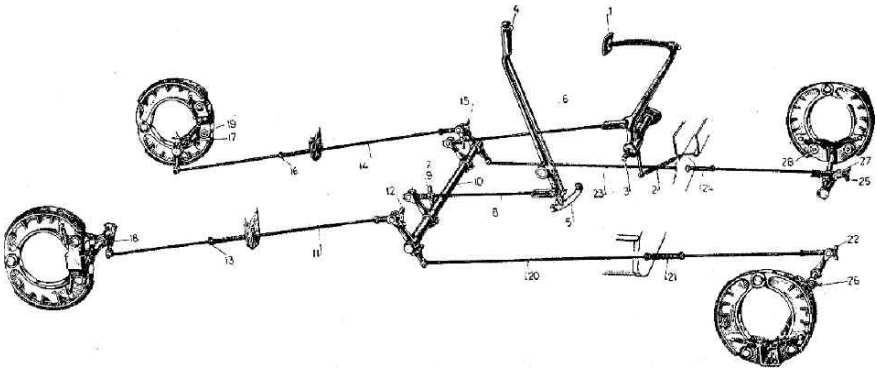


FIGURE 1.37. Mechanical control for a four wheels braking system; the articulation points between levers and tie rods had to stay on the center of instantaneous rotation of the axles during the suspension stroke, located near the fixed eyes of the leaf springs.

consisted in rounding the cam in such a way that contact points would always be on the king-pin axis.

Thus brake control was made insensitive to steering; this patent was overtaken by other mechanical solutions and four wheel braking was state of the art at the end of the 1920s.

What we said about steering motion can be also applied to the suspension motion of both axles.

The distance between the articulation points of tie rods could not be allowed to change over a certain threshold during the suspension stroke; if this happened the brake could be actuated by road bumps or the tie rods could be broken by excessive pull.

For this reason, as we can see in Fig. 1.37, the articulations points between the levers and tie beams had to stay on the center of instantaneous rotation of the axles, located near the fixed eyes of the leaf springs during the suspension stroke.

The tie rods must have adjustable length, to adjust for wear of the shoes; notice also that the distribution of braking force is influenced by the length of the rods.

These problems were solved by the application of hydraulic controls, which spread during the 1930s. These were proposed by Lockheed and introduced on a car for the first time by Duesemberg in 1921.

Another important development of braking systems was the application of auxiliary power, which increased the driver's control over the hydraulic lines; this device was introduced using engine vacuum as its energy source. It was applied for the first time on a Ford car in 1954 and allowed even the weakest drivers to brake efficiently.

As early as 1910 Rolls-Royce applied on its cars a completely mechanical power brake, shown in Fig. 1.38.

www.TechnicalBooksPDF.com

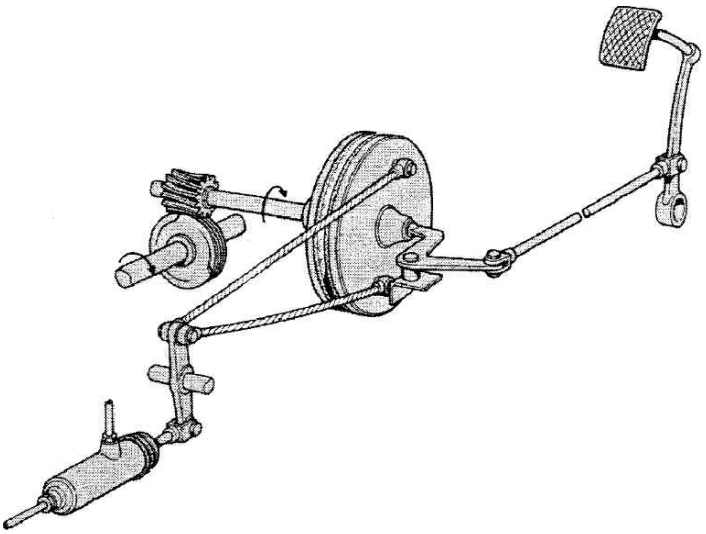


FIGURE 1.38. A mechanical power brake adopted by Rolls-Royce; the necessary energy derives from the propeller shaft and is transferred to the braking system by a clutch, operated by the brake pedal.

The mechanical energy necessary to operate the brakes is derived from the engine, transferred to the braking system through a clutch that slides in a controlled way managed by the brake pedal. The energy from the propeller shaft also contributes to slowing the car.

1.6 CHASSIS FRAME

To speak of the structure of the chassis we must say something about car body manufacturers.

At the dawn of the automotive industry, people who dedicated their financial and intellectual efforts to this product preferred, understandably, to concentrate on the most difficult problems of car development, particularly the engine, transmission, suspension and steering system. In other words, they became *chassis* manufacturers, according to French term.

The body, according to technical rules known at the end of the XIX century, was not considered a critical technology. Indeed it could be totally imported from coaches and carriages; many wheelwrights also became car body manufacturers.

This fact caused an initial separation of these two kinds of industry.

Chassis manufacturers used primarily metallic materials and their plants were able to cast, stamp and machine; due to the precision of couplings they worked according to drawings and produced small product runs.

Body manufacturers used primarily wooden structures and worked as carpenters, using fixtures and tools to produce single pieces, often without drawings.

www.TechnicalBooksPDF.com

This tradition drew on the fact that wood was much more suitable than metal, for obtaining curved shapes that already followed elaborate aesthetic rules.

Painting, vital for surface protection and appearance, also favoured wood, considering existing oil paints and varnishes. This justified a complete separation of chassis and body fabrication to avoid damage during the complicated assembly of mechanical parts.

It should be remembered that the complete treatment of a body required more than 400 h of work; to this time the drying time of any layer had to be added.

The chassis had, therefore, an important technological value required by industry. It allowed the car manufacturer to produce a finished product that could be delivered to the body works, relieving the car manufacturer of any further responsibility; it could also be driven for testing and to demonstrate its performance to the customer.

The final customer bought, therefore, a chassis that was handed off to a coach builder for completion. In some cases coach builders bought chassis on their own and sold finished cars. Some car manufacturers had permanent agreement with coach builders to produce and sell complete cars.

This was the situation in Europe at the beginning of the First World War.

After the war mass production started and many manufacturers introduced internal body shops. This fact produced a gradual transition from wood to steel, with intermediate solutions we could define as hybrid, using a wooden skeleton and external panels made of wood or partially of steel sheets.

Synthetic enamels shortened the painting time by an order of magnitude and made a better integration of body and chassis fabrication possible.

The organization of the production cycle did not change greatly because of tradition, and because customers still requested special bodies that needed to be assembled on a finished chassis.

The integration of chassis and body manufacturing was developed primarily in the United States and imitated in Europe by major manufacturers, beginning in the 1920s.

The frame that is the bearing structure of the chassis had to carry all mechanical components, the complete body and payload. In addition the frame was the assembly support for all chassis components, including the engine, needed to obtain a rational organization of the fabrication process of the car.

The frames of early cars were made either of wood or steel; this last solution was more likely applied. Figure 1.39 shows an example of a complete chassis of an early XX century car, where the frame can be easily identified.

The frame is made in a *ladder structure*, showing two longitudinal *side beams* with a particular shape.

In front the distance between the two beams is reduced to allow for the necessary space to steer the front axle wheels, while the rear side is larger to better install the car body; in this area the distance between side beams is determined by the track and the dimensions of the transmission.

www.TechnicalBooksPDF.com

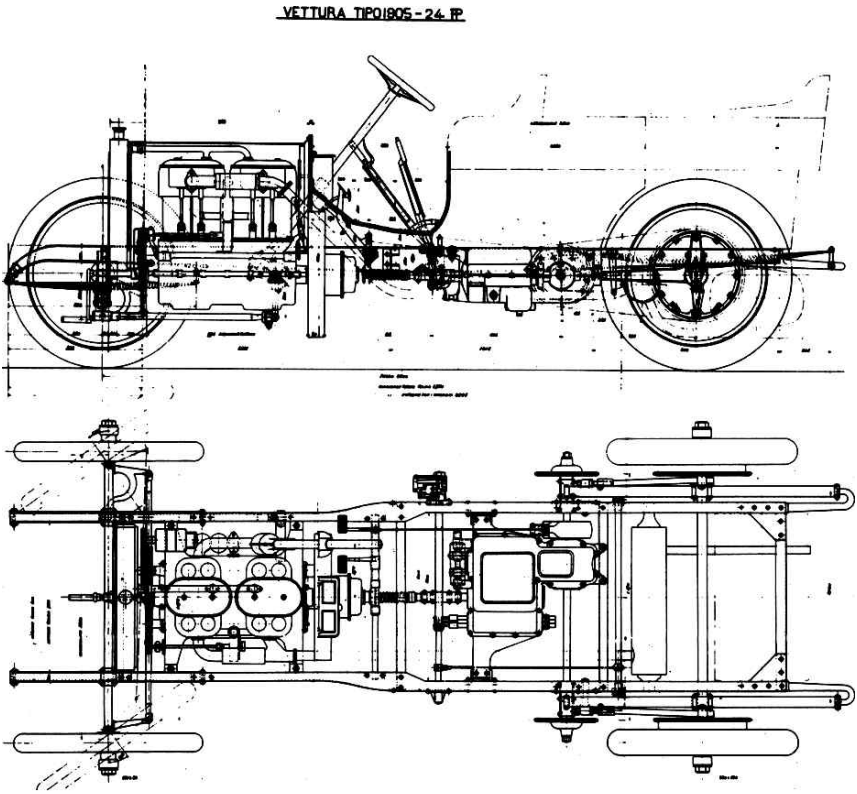


FIGURE 1.39. Complete chassis of a Fiat car of 1905. The chassis structure made of a ladder frame is similar to that of many cars of that era.

The front and rear ends are often curved on the side to better match the shape of the leaf springs to allow for the suspension compression stroke. The curvature of later cars was increased to reduce the height of the floor between the wheels.

The side beams are connected by a number of *cross members*, shaping the ladder structure.

The side beams and cross members are bolted in our example, but in other cases the were heat riveted and, later, lap welded.

This kind of structure has been used for years and is still found on many industrial vehicles; the choice of bolts or rivets instead of welding was imposed by the dimensions of the structure, The need being to reduce unacceptable residual stresses after welding, that might be caused by the quantity of applied heat.

The torsional stiffness of this kind of frame is not particularly high, because it is determined by the cross section of the side beams, which are almost always of the open type to make the assembly with cross members and chassis components easier.

www.TechnicalBooksPDF.com

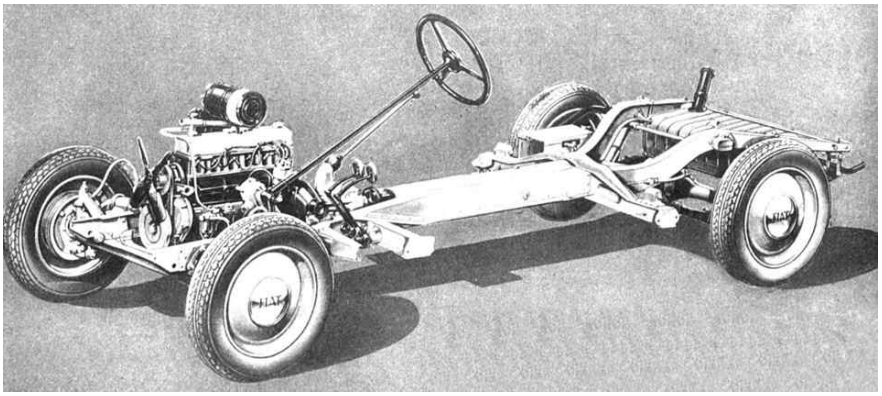


FIGURE 1.40. Central beam frame structure of Fiat 1500 of 1935; the structure is made of steel sheet pieces stamped and riveted to each other.

The bending stiffness is low because of the limitations on the height of the side beams.

Chassis frame deformations became an important issue because of their effect on the body; the problem became more evident as maximum speeds increased.

The body was made by a wooden skeleton that was quite stiff after fabrication, because of the carpentry joints that could be applied. These joints, however, could not fulfil the task of making the entire car stiff.

Chassis deformations applied to a stiffer body resulted in loosening or ruptures of the body skeleton joints; such deformations created play between body parts and, therefore, annoying squeaks and rattles.

Body joints were perfected by abandoning carpentry solutions in favour of more flexible metal joints that avoided contact between parts; to this purpose fabric covered structures were also developed, according to Weymann's patents. These came from the aeronautical technology of that era, which allowed for a very flexible but quiet body.

The final solution came through the use of steel sheets, at least for outside body panels; on one hand the increased flexibility better matched chassis deformations; on the other hand it could contribute to stiffness of the entire structure, being able to carry an increased stress.

The chassis frames of some cars from the 1930s changed in aspect and showed more elaborate designs, as can be seen in Fig. 1.40.

We can see in this figure that the frame elements are made with stamped parts. Side and cross members, because of their complicate shape, are deep-drafted using dedicated stamps; in the previous example, in Fig. 1.39, the same components were cut from profiled steel bars and bent with open stamps.

The advantage of this innovation is better control over structural weight, because every part of the frame is designed according to local stress; the disadvantage is obviously the higher amount of dedicated tool investments.

www.TechnicalBooksPDF.com

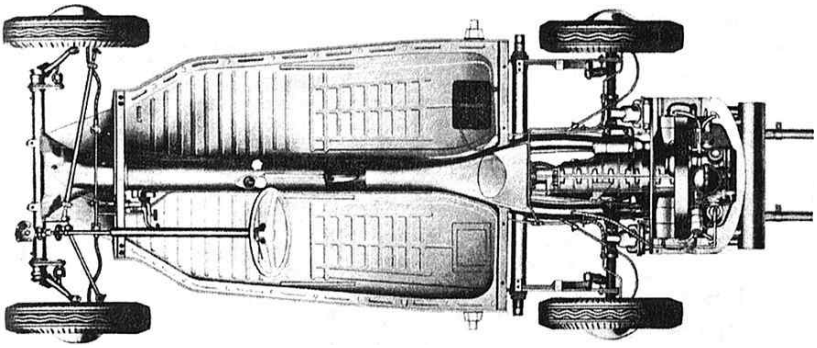


FIGURE 1.41. Platform chassis structure of Volkswagen Beetle of 1936. This structure is separated from the body during the assembly of the chassis, but after body assembly shapes up as a closed structure, whose torsional stiffness can be compared to that of a unitized body.

Instead of the already shown ladder frame, we now have a cross frame where side beams are shaped to build an X; the cross members, instead of connecting side members, shape up cantilever beams along the floor and the body. The two side members are joined together in the middle, building up a tubular structure of greater torsional stiffness; the scheme of the frame allows the floor to be in a lower position, with benefits for the car height.

The front cross member is a cast element bolted to the frame, able to receive the independent suspension shown in Fig. 1.15.

Most joints are riveted.

This kind of chassis is connected to the body through robust bolts. The body was made by stamped steel panels imposed upon a wooden skeleton; the assembly is not particularly stiff but should contribute to the total torsional stiffness.

A further step appears in the Volkswagen Beetle, almost contemporary with the cars previously shown, but produced for a longer time; the chassis structure is shown in Fig. 1.41.

The Beetle's chassis structure involves a central beam to which floor and front and rear cross members are welded to support the front suspension (shown in Fig. 1.16), rear suspension and power train.

The floor is bolted to an integral full steel sheet body that is welded completely; this solution cannot be called a unitized body, but its structural behavior is comparable when assembly is completed

This kind of chassis structure, called a chassis *platform*, was still in use in many cars until the 1950s.

It shows the advantage of being compatible with the assembly process already described; this solution was obviously suitable for producing different car variants on the same chassis.

The unitized body was thought to obtain optimum structural performance with a reduced weight; the considerable distance, existing in a sedan car, between

www.TechnicalBooksPDF.com

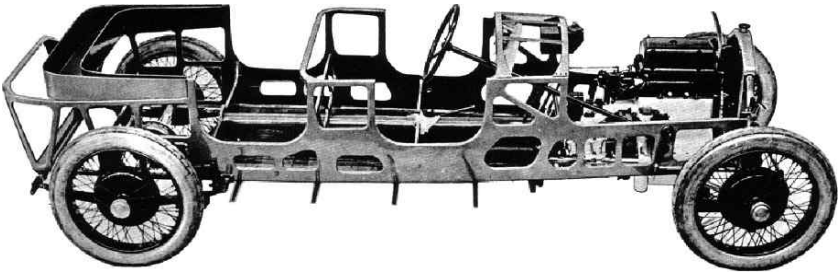


FIGURE 1.42. Complete chassis with parts of the unitized body of the Lancia Lambda of 1922. This picture is only for illustration; in fact, mechanical components were assembled to the structure only after the completion and painting of the body.

chassis side members and longitudinal roof reinforcements allows a very stiff structure, if these elements are connected with vertical members (*pillars*) sufficiently stiff and well joined to longitudinal and cross members.

Body panels, if shaped for this purpose, can also contribute to structural stiffness by limiting angular deformations of pillars, side beams and cross members.

The first car built using this method was the Lancia Lambda of 1922, whose chassis is shown in Fig. 1.42.

The body is made by riveted panels deployed so as to build up boxed areas; many parts of the structure are supported by side beams, integrating seats, fire wall and trunk and contributing impressively to the total stiffness.

The front suspension is connected by a gate structure surrounding the cooling radiator, as already shown in Fig. 1.13.

We are not able to comment quantitatively on the structural performance of this solution, but the result in the volumes layout of this car is impressive, allowing the modern and slender shape characteristic of this car as compared with its contemporaries.

This structure could fit a convertible body; sedans did not take advantage of it, featuring a Weymann's semi hard top instead.

The crucial step was taken about fifteen years later when spot welding was developed and deep stamped steel sheets became available; it should be noticed that every cut and bend of the Lambda structure is made by open and general purpose stamping tools.

One of the first examples in Europe of this technology, developed by Budd in the United States, was the Fiat 1400 of 1950, represented in Fig. 1.43.

All body panels are shaped in such a way as to build boxed beams in the junction areas; these boxed beams create a spatial frame that supports all structural tasks.

For example the spacial frame supporting the power train is made by two members shaped like a U (on the upper side) that along with the sides of the wheel wells build up closed section beams.

www.TechnicalBooksPDF.com

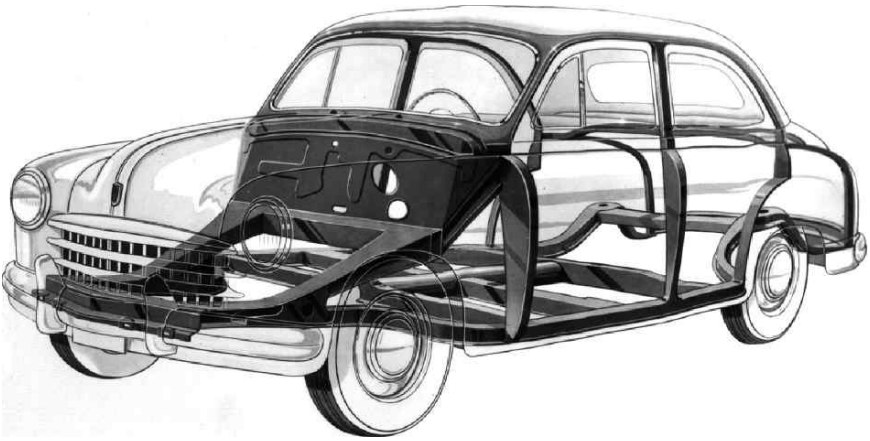


FIGURE 1.43. Phantom view of the unitized body of the Fiat 1400 of 1950; darkened areas reveal the body structure. The chassis as physical subassembly no longer exists and all components (power train, suspension, steering system, etc.) are bolted to the body.

The chassis, as a physical subassembly of certain mechanical components, no longer exists, its separation from the car impossible; nor can its structural function be identified. Nevertheless the structural function of the body is evident in the darkened elements of the phantom view.

The production process is also completely different; the body is stamped, welded and painted and chassis components grouped by small sub assemblies are assembled only afterwards.

The role of body manufacturers has been completely revolutionized.

www.TechnicalBooksPDF.com

2
WHEELS AND TIRES

2.1 DESCRIPTION

Vehicle wheels have essentially two functions:

- To support the weight of the vehicle by exchanging vertical forces with the road surface
- To exchange with the road surface longitudinal and side forces, which are able to move the vehicle and control its path

As we have seen in the previous chapter, the first function was already present on antique wheels that had to carry weight but were only occasionally demanded to brake the vehicle or withstand significant side forces.

With increasing speed the bearing capacity for side and longitudinal forces becomes more and more important.

Road vehicle wheels include two elements: the rim and the tire.

2.1.1 Rim characteristics

A unique characteristic of a rim is that it allows fast and simple substitution of the tire. A mean average tire life should range between 30,000 and 60,000 km, according to the weight of the vehicle and how it is used. In addition, tires can be punctured and require immediate substitution.

These facts justify the particular shape of the rim and the fact that its dimensions are standardized in order to be exchangeable with those of other manufacturers.

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

www.TechnicalBooksPDF.com

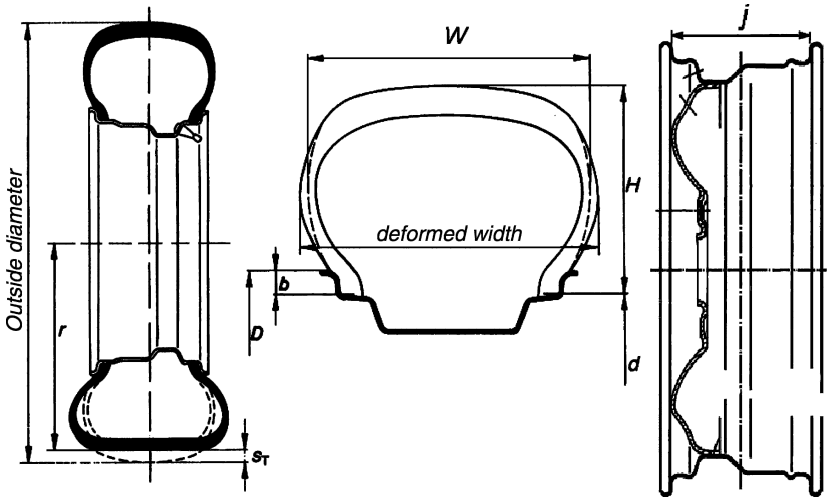


FIGURE 2.1. Schematic section of a complete wheel including rim and tire and their primary dimensions.

The wheel is made up of a *disc* and a *flange*; these are shown on the right side of Fig. 2.1; in the same figure can be seen the undeformed section of the tire (on the upper side) and the deformed section (on the lower side) of the contact.

Disc and flange are usually integral; the wheel is fixed to its hub with bolts.

Flanges show a particular shape suitable for keeping the tire in place. The flange is characterized by the *diameter* d and shows a dropped center useful for tire assembly and disassembly; in fact the bead cable of the tire is quite inextensible and must be set initially into the dropped center from one side, crossing the flange for final installation.

The j dimension between the two flanks of the flange determines the mounting width of the rim; dimensions d and j are usually measured in inches and must be identical to the dimensions of the tire.

The wheel size measurement is usually written as in the following example:

$$5\frac{1}{2}\ J\ \times\ 15,$$

to show a rim with a j dimension of $5\frac{1}{2}$" (139.7 mm) and a d diameter of 15" (381 mm); the letter J shows the type of adopted rim profile, almost always of this type.

Groups of letters can be added to this statement to show the *hump* type of the flange; the hump is the shape of the area matching the internal diameter of the tire. A particular tapered shape is used to improve the air seal for tubeless tires or to avoid sudden tire removal after puncturing. No letter stands for a conical rim for tube tires.

www.TechnicalBooksPDF.com

Rims can be made of stamped and welded steel or cast aluminum or magnesium; this last alternative, initially developed for weight reduction, is also justified by aesthetic reasons.

The rim disc is characterized by windows suitable for improving brake cooling and by a central hole able to sustain the wheel before the bolts are tightened; the wheels fit the disc in a tapered area to increase disc elasticity. This has the function of reducing the statistical pull dispersion after tightening.

Rim discs usually display a bell shape intended to give the correct position to the symmetry plane of the tire (*equator* plane) with respect to the car and to allow the necessary space for hub and brake installation.

Today many cars can be fitted with tires of different size, according to the engine power of the model and the desire of the customer; usually higher performance versions have a wider tire with the same rolling radius. As a consequence the aspect ratio is reduced to increase cornering stiffness, though with compromising effect on comfort.

The disc shape of the rim is usually such as to maintain the outside profile of the tire in the same position.

2.1.2 Tire characteristics

The rigid structure of the wheel is surrounded by a flexible element made up of the tire and its *tube*, which are used to maintain inflation pressure. This last can be absent, as in *tubeless* tires, where the tire is hermetically fitted to the wheel; this more costly solution is today preferred because of enhanced safety, because punctured tubeless tires are slower to loose their pressure.

The tire is a complex composite structure, made up of many layers of rubberized fabric (*plies*) with reinforcements cords; the orientation of warp and weave and of reinforcement cords gives a tire its unique mechanical design characteristics.

The angle between the direction of the cords with respect to the circumferential direction of the tires is usually referred to as the *crown angle*. As a general rule, plies with a low value of crown angle enhance the handling characteristics of the vehicle, while those with a high value, up to 90 deg, of crown angle enhance ride comfort.

Each tire is designated by a group of letters and numbers, as in the following example:

$$175/65\ R\ 14\ 82\ T.$$

We explain the meaning of this designation with reference to Fig. 2.1.

- The first figure (175) shows the dimension W, usually measured in mm; it is separated by a bar from the following figure. Because the tire is a deformable structure, this measure must be referred to an undeformed situation with correct inflation pressure and no load applied.

www.TechnicalBooksPDF.com

TABLE 2.1. Letters showing the maximum allowed speed of a tire.

Speed (km/h)	80	130	150	160	170	180	190	210	240	270
Letter	F	M	P	Q	R	S	T	H	V	W

- The second figure (65) refers to the *aspect ratio* of the tire, given by the ratio H/W between the radial height and the width; aspect ratios are usually expressed as a percentage: In the given example H is 65% of W. If this figure is omitted it should be assumed to be 80%, long a standardized value.

- The following letter shows the type of tire plies; R stands for radial. The designation is otherwise omitted.

- The third figure shows the rim diameter in inches.

- The fourth figure is the *load factor*, which determines the allowed vertical load at an assigned inflation pressure; this figure is the row order of a standardization table where maximum allowed vertical loads are shown as a function of inflation pressure. This figure therefore has no physical meaning.

- The final letter indicates the maximum allowed speed for the tire, as shown in Table 2.1.

Other standard designation, no longer current, survive in older publications; for example:

$$P\ 225/50\ SR\ 15.$$

In this case, the first letter stands for vehicle category (P is for cars), the first two figures indicate tire width in millimeters and aspect ratio, the following letters shows the maximum allowed speed and the next letter the tire type (R for radial, B for belted, D for cross-ply); the last figure is the rim diameter in inches.

Independent of their application, tires can be made according to different fundamental schemes; those regarding the tire *carcass*, i.e. the organization of plies, can be classified according to two categories: *cross* or *conventional ply* and *radial ply*. A radial ply tire is shown in Fig. 2.2; *belted tires* could be considered an intermediate category.

On cross ply tires reinforcement cords cross the equator plane with a crown angle of $35 \div 40$ deg; on radial tires some plies running perpendicular to the circumferential direction (carcass plies, with a crown angle of 90 deg) are surrounded by other plies (belt plies) with a much smaller angle, in the range of 15 deg. On cross ply tires, plies extend from bead to bead, while on radial tires belt plies are limited to the region that comes into contact with the ground, while the carcass plies hold the sides of the tire into shape.

www.TechnicalBooksPDF.com

FIGURE 2.2. Radial tire structure: the tires include plies with circumferential cords (belt) and plies with radial cords (flanks) and two reinforcement cables in the bead for that part of the tire contacting the rim.

The structure of conventional tires is less vulnerable in the flanks but more flexible in the belt region; radial tires are, on the other hand, very flexible and vulnerable in the flanks, being stiffer in the belt region: This arrangement improves cornering and comfort with a penalty in flank vulnerability (shocks against sidewalk steps, etc.).

Because of their superior characteristics, radial tires are the most common type; conventional tires are in use for more demanding applications such as off-road driving.

The primary function of tires is to distribute the vertical load on a surface which should be large enough to absorb road irregularities. It is vital that the flexibility along different directions be distributed in a specific way: on radial tires, for example, vertical deformation is high because of the high deformation of the sides, while circumferential deformation is low, because of the almost straight direction of the belt cords. Superior comfort and high capability to generate side forces are a direct consequence of this flexibility distribution.

Particularly relevant to tire behavior is the *tread* of the tire, essentially made of vulcanized filled rubber; it is the contact surface with the ground and determines friction at the tire-ground interface. Treads feature shapes in different designs; circumferential and transversal grooves have the essential purpose of making it easier for water to drain away from the contact area when the road is

www.TechnicalBooksPDF.com

wet. They also improve friction on incoherent soils: Thinner transversal grooves improve friction on slippery ground. If the vehicle drove only on well paved and dry roads, such grooves would be unnecessary, as in the case of *slick* tires.

2.1.3 Wheel reference system

While studying forces between tire and ground, it is useful to use the wheel reference system $X'Y'Z'$ shown in Fig. 2.3. The origin of the system is the center of the contact patch, defined as the theoretical contact point between the undeformed equatorial plane and the ground; the X' axis is the intersection of this plane with the ground and its positive direction is the same as for vehicle speed. The Z' axis is perpendicular to the ground[1] and, as a consequence the Y' axis lies on the ground and points to the left.

The total force is supposed to be applied at the center of contact. It can be seen as three components acting along the X', Y' and Z' axis, as *longitudinal force* F_x, *lateral force* F_y and *vertical force* F_z. In the same way the total moment

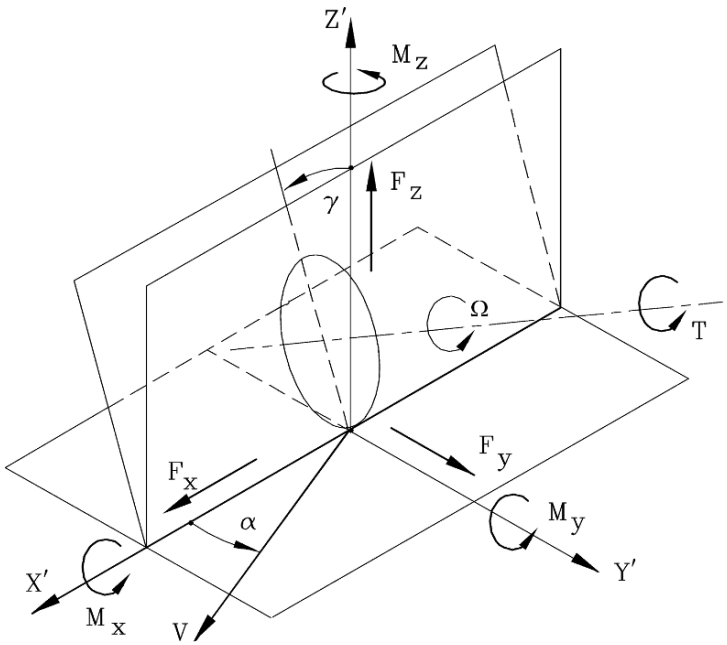


FIGURE 2.3. Reference system used to study forces exchanged between tire and ground. Definition of positive direction of forces, moments and side slip angle.

[1]The terminology we are going to use is that suggested by SAE (Society of Automotive Engineers) in its document J670e Vehicle Dynamics Terminology, published in 1952 and revised in July 1976. Nevertheless in some cases the Authors have decided not to follow some suggestions, as in the positive direction of the Z' axis, which in this document points downwards.

www.TechnicalBooksPDF.com

can be broken down into three components: the *overturning moment* M_x, the *rolling resistance moment* M_y and the *self aligning moment* M_z. A *wheel torque* T can be applied by the vehicle along the tire rotation axis.

The angle between the $X'Z'$ plane and the direction of the wheel hub motion is called *wheel sideslip angle* α, while the angle between the $X'Z'$ plane and the wheel equatorial plane is called *inclination* or *camber angle* γ. Positive directions for angles are shown in the same figure; in particular, the inclination angle is positive when the highest part of the wheel is oriented to the right of the driver (in other words, the left wheel has a positive inclination angle if its top is points to the car interior). While the inclination angle is stated with reference to the road, the camber angle is usually stated with reference to the car and has positive directions that are opposite for the two sides of the vehicle.

In this book we will assume that inclination and camber angle coincide.[2]

2.2 TIRE OPERATION

While explaining tire operation, we must identify two different situations:

- Prepared ground, when the tire is in contact with paved or concrete surfaces
- Unprepared ground, when the tire is in contact with natural surfaces or dirt roads

The first situation is typical of on-road driving, while the second is typical of driving off-road; the physical phenomenon discriminating the two is that ground deformations should be considered or neglected. Ground deformation is neglected on dry paved roads.

2.2.1 On-road driving

Two different aspects of on-road driving are to be considered:

- The *adhesion* between rubber and ground; because of this phenomenon, tires can exchange with the ground forces that are contained in the contact plane.
- The *elasticity* of the tire structure, which gives to the tire certain absorption capabilities as far as road irregularities are concerned. These are the primary reason that tires slip in two directions of the area of contact when longitudinal and lateral forces are applied.

[2]This is a further deviation from SAE recommendations.

www.TechnicalBooksPDF.com

Rubber-ground adhesion

We define adhesion as the result of physical phenomena that allow a specimen of rubber set on the ground and pressed with a certain vertical force to withstand forces contained in the ground plane, without any relative motion.

Adhesion is caused by two different phenomena that have similar results:

- *Physical adhesion* and
- *Local deformation*

Let us consider the physical adhesion phenomenon by looking at a piece of rubber pressed onto a completely smooth and rigid ground. Rubber molecules and ground molecules apply certain attraction forces; these forces refer to distances in the range between $10 \div 10,000$ Å (equivalent to $0.001 \div 0.01$ μm). Impurities present between the two contact surfaces limit the adhesion phenomenon to a certain number of adhesion sites on both surfaces.

Adhesion forces will develop between adhesive sites. If we now imagine applying a lateral force to our rubber specimen, we will see that it is balanced by the adhesion force.

This force will be maintained until a certain distance is reached between the adhesive sites and will then become zero, as the distance exceeds a certain value; when the adhesion force is destroyed, molecular layers are instantaneously freed from stress and begin oscillating in the material. Because of internal material damping this process will dissipate a certain amount of energy; the dissipated energy divided by the displacement between the rubber specimen and ground will result in a reaction force opposing this displacement.

Adhesion force is therefore controlled by the following parameters:

- Surface energy of contacting materials.
- Damping properties of those materials, where rubber plays the most important role; these properties are controlled by temperatures and bound to relative speed.
- Deformation of contacting surfaces, because of lateral forces which can also cause instability as in the case of *stick and slip*.

Adhesion force can be measured by pressing a rubber specimen onto a perfectly polished mirror-like surface.

The force should appear because adhesion is bound to material damping properties.

Local deformations are, instead, caused by road irregularities; if we exclude adhesion, as could happen if we put our rubber specimen on a perfectly lubricated but rough ground, we will have areas where the rubber deforms and afterwards looses its deformation.

Again, mechanical work dissipated by damping will build up forces along the contact surface.

www.TechnicalBooksPDF.com

In addition, friction and damping are bound to each other.

Local deformation forces can be measured by having a rubber carpet specimen slide on a low resistance plane, such as a plane equipped with rollers or spheres that are free to rotate.

Adhesion force comprises about 70% of the total friction force when rubber is on dry paved road.

These phenomena change their mechanism radically on wet surfaces; we again distinguish three fundamental cases.

- Water layer thickness is high enough to establish a permanent lubricated meatus between tire and ground (*aquaplaning*); in this case adhesion and local deformations cannot take place and tangential forces can be calculated, according to lubrication theory, from liquid viscosity.
- Water layer thickness is insufficient to establish permanent lubrication, but sufficient to preclude adhesion forces; in this case local deformation forces can still occur if the ground is rough.
- Water is completely removed from the contact area. The behavior of rubber is then as was explained.

This last case is what we would like to obtain by means of tire grooves and draining pavement.

In real situations it can happen that the first part of the patch will receive a lubricated hydrodynamic lift, while the areas behind are in intermediate or dry conditions.

We define *friction coefficient* μ as the ratio between the lateral force and the vertical force; as forces, it can be broken down into a longitudinal friction coefficient using the ratio:

$$\mu_x = \frac{F_x}{F_z} \tag{2.1}$$

between forces along the X' axis, and a vertical force along the Z' axis and longitudinal friction coefficient, using the ratio:

$$\mu_y = \frac{F_y}{F_z} \tag{2.2}$$

between the force along the Y' axis and the vertical force along the Z' axis.

Friction coefficient can be defined globally, as we did, or locally as the ratio between tangential and vertical pressures.

Elastic behavior

The tire can be compared to an elastic deforming structure with internal damping.

Without inflation pressure tire rigidity will be modest and inadequate for practical use. By neglecting the mechanical contribution of cloths and cords, the

www.TechnicalBooksPDF.com

tire could be thought of as a perfect membrane with constant internal pressure, on which a certain vertical load is applied on a part of its structure.

This membrane could withstand vertical load only by its internal pressure; the tire surface should flatten for an area big enough to generate a pressure-area product able to equal the vertical force. This surface is the *contact patch* of the tire.

The contact patch area decreases if the inflation pressure increases.

If the tire is rolling previously deformed patch areas must return to their normal shape, while new material must be flattened so as to withstand the vertical force.

It can be imagined that this process implies energy dissipation because of the damping properties of rubber. These damping properties are caused not only by internal material damping, but also by the relative motion that must take place between cloths, cords and rubber, to allow for local bending deformation.

If the ground is not rigid, anelastic ground deformation plays a role in determining roll resistance.

The membrane model easily explains why lateral and longitudinal forces can induce local deformation of the contact patch; what is more difficult to understand is what is happening when lateral and longitudinal forces are applied to a rolling wheel.

To predict, although only in a qualitative way, what will occur we will use a simpler model, represented in Fig. 2.4.

The tire is reduced to a discrete number of springs, featuring vertical, lateral and longitudinal flexibility and bearing a specimen of tire rubber on their tip.

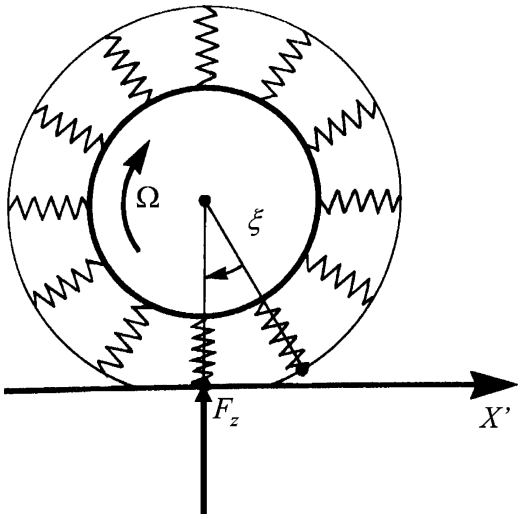


FIGURE 2.4. "Brush" simplified model of a tire; the tire is reduced to a discrete number of springs, featuring vertical, lateral and longitudinal flexibility and bearing a specimen of tire rubber on their tip.

www.TechnicalBooksPDF.com

Let us assume, for the sake of simplicity, that the number of springs is such that there is a position where one only of them contacts the ground; they are fitted on a rigid rim with a uniform angular displacement ξ.

The spring tips are free to move, according to applied forces, with no connection to nearby springs. All simplifications can be removed without affecting the nature of the results we are going to study. We will call this model, extremely simplified, the *brush model*.

Under the application of the vertical force, a contacting spring will assume a certain radial deformation, as in a real tire; if we associate with the spring a suitable damping coefficient, the model is also able to simulate rolling resistance proportional to the mechanical work dissipated by any contacting spring as it leaves the contact patch because of rolling.

In addition, rolling radius, because of the applied vertical force, will assume a value smaller than the unloaded value.

Let us assume that the tire undergoes a longitudinal force F_x in the driving direction (middle illustration of Fig. 2.5) or the braking direction (right illustration of Fig. 2.5); according to the proposed model an angular deformation ζ, proportional to the applied force, will be applied to the contacting spring and the tire.

It is easy to understand that if the tire has a rolling speed Ω, each contacting spring will return to a straight position when leaving the contact patch, while a new spring will assume the angular displacement ζ. To the initial tire speed Ω a new deformation speed will be added, equal to the ratio between the angular deformation ζ and the time ξ/Ω needed to put the following spring in contact with the ground.

Under the action of a driving force the wheel should roll faster than when rolling free of force; on the other hand, if the wheel is braking, it should roll slower than when rolling free.

If our model was accurate speed variation should be proportional to applied longitudinal force; this speed variation is called *longitudinal slip speed*.

When the braking or driving force goes over the friction limit, a new additional speed should be generated depending only on mass characteristics; over this value longitudinal forces can no longer increase.

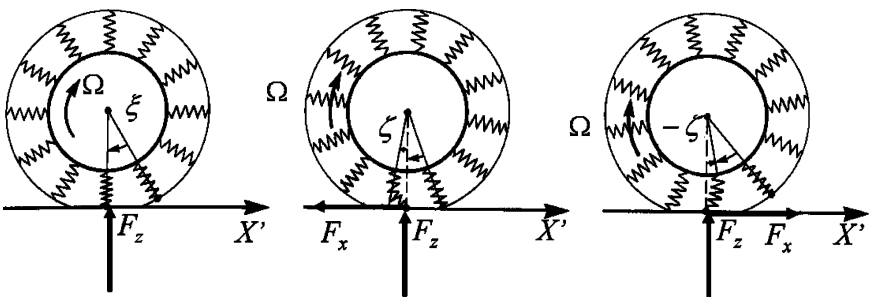


FIGURE 2.5. Brush model: static representation of radial and circumferential deformations of a tire, with contemporary application of a vertical, braking or driving force.

www.TechnicalBooksPDF.com

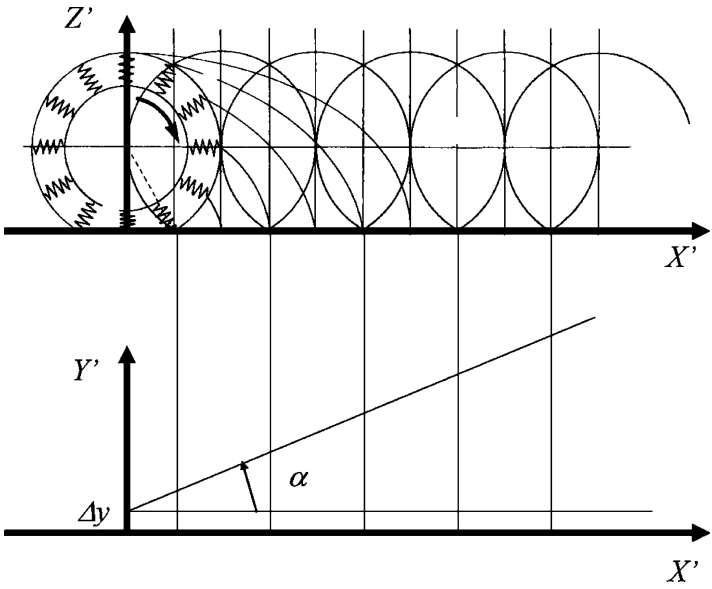


FIGURE 2.6. Brush model; graphic representation of a rolling tire with contemporary application of vertical and longitudinal force.

Let us assume, finally, that our tire undergoes a vertical and lateral force due, for example, to a centrifugal force applied to the vehicle in a turn. In this case, the spring in contact with the ground will assume a lateral displacement in the Y' direction. This phenomenon is shown in Fig. 2.6.

If the wheel is rolling, each spring will contact the ground as shown in the upper part of this figure, here seen relative to the side view on a certain point of the X' axis.

The initial contact point will receive a lateral displacement Δy, shown in the lower part of the same figure; this displacement again is proportional to the lateral force. When the next spring contacts the ground it will assume the same displacement, which will be added to the wheel hub. As a result the wheel hub will receive a lateral motion whose speed will be given by the ratio between this displacement and the time needed to place the next spring in contact with the ground, that is ξ/Ω.

If the model is accurate, because deformation is proportional to the force, we would expect the wheel hub to deviate in its path by an angle α, proportional to the lateral force; this deviation is called *sideslip angle*. It is more accurately called tire sideslip angle, not to be confused with the vehicle sideslip angle or attitude angle, we will explain later on.

When lateral force exceeds the friction value of rubber, the added speed will be controlled only by masses and the side force should have no further increase.

The brush model has only given us the possibility to understand and predict in an approximate way the concepts of longitudinal slip and sideslip.

www.TechnicalBooksPDF.com

We will examine longitudinal slip and sideslip, in a better detail, by means of actual experimental results and their numerical interpretation; we do not suggest to use the above models for any quantitative predictions.

2.2.2 Off-road driving

In off-road driving, besides the phenomena we have described (which remain unchanged also in this case), ground deformations should be taken into account. They could predominate in certain situations.

Some vehicles are designed to be driven not only on paved roads but also on natural ground, on country and forest roads with surfaces only roughly levelled. Prevailing factors limiting vehicle mobility are the deformation of the surface and the shape of the ground, which can interfere with the mechanical parts of the chassis.

We will discuss ground deformation in the following paragraphs. On a level road, mechanical behavior caused by untreated ground can have a severe impact on vehicle motion. The shape of the surface has a purely geometric impact on chassis design, concerning:

- Wheelbase
- Track
- Wheel diameter
- Available suspension stroke
- Clearance of the chassis from the ground and
- Attack angles

We will refer to natural soils and to fresh snow, which for certain reasons shows a similar behavior.

Natural soils are the final result of wind and water erosion on the rocks from which they have originated. They are made of tamped down powders.

The mechanical properties of soil are determined by the solid incoherent particles that are its main component; The primary characteristics of these particles are their granulometry, their apparent density and the content of water or *humidity*. We distinguish between *apparent* and *real density* to account for the space separating contiguous particles. The first is defined by the ratio between weight and volume of an average lump cut from the ground and measured as is; the second represents the limit of compression possible for this lump in order to eliminate any space between particles; in other words it is the average density of single particles.

The forces that can be exchanged by nearby particles and that oppose deformations caused by externally applied forces, are cohesive forces that usually

www.TechnicalBooksPDF.com

increase as particle size, or *granulometry*, decreases; in fact, if granulometry decreases the contact surface between nearby particles increases.

The most significant contribution to soil deformation is produced by particles sliding along their contact surface; the higher the cohesion force the lower the deformation. Cohesion forces are also influenced by the presence of water trapped between nearby particles.

On the other hand, a significant contributor to limiting particle displacement and consequent soil deformation is the friction force at the point of contact between particles; the friction coefficient is also influenced by the presence of water.

At low levels of humidity, soils feature a high value of mechanical resistance that decreases as the quantity of trapped water increases.

The *liquid limit* is reached when water is so abundant as to completely eliminate the cohesive effect between particles and therefore the shear resistance of the soil; this limit is expressed by the quantity of water divided by the dry residual.

The *plastic limit* is reached when the soil looses its capability of being shaped, being too brittle. This limit is expressed by the quantity of trapped water divided by the dry residual.

The difference between the two limits is called the *plasticity index*, while the ratio between the actual water content of a given soil and the quantity related to the plastic limit is called the *relative content of water* of the soil.

For our purposes we propose the following classification of soil types.

Cohesive soils

Cohesive soils contain variable quantities of clay, mixed with very fine granules of sedimentary rock, with a mean diameter between $0.05 \div 0.10$ mm. They can be classified as *lightly clayey*, *clayey soils* and *clays*, according to the quantity of clay. In these soils cohesion is particularly relevant and is dominated by relative humidity; for this purpose solid, plastic and fluid states are identified.

Relative humidity is therefore an important parameter in describing the mechanical properties of cohesive soils; these soils are stable only if relative humidity is in the range of $0.4 \div 0.65$; over these values mechanical properties deteriorate severely. A peculiar characteristic of these soils is that of draining water with great difficulty; vehicle mobility will become particularly difficult during rain season and times of ice melt.

Cohesive soils occupy a large part of the earth surface.

Sandy soils

In sandy soils particles are larger and clay is practically unknown; a possible classification of these soils would take into account particle diameter and apparent density. The mechanical resistance of sandy soils to deformation initially increases as water content increases, while decreasing rapidly at higher contents;

www.TechnicalBooksPDF.com

unlike clayey soils, sandy soils have a very good draining capacity: They are widespread on sea and river shores and in some deserts.

Swampy soils

Swampy soils are characterized by a high content of peat, where most diffused constituents are residuals of animal and vegetal organic matter, at the end of an anaerobic decomposition process; they can be classified as *solid swampy soils*, when peat lies directly on a rock layer, *saprofelic soils*, when the peat lies over a layer of materials in decomposition and *floating soils* when peat floats over the water.

These soils are widely found in Central Europe and Siberia.

Snowy soils

Snow is particularly abundant during the winter season of the Northern Hemisphere. Vehicle mobility on snow depends primarily on snow layer thickness, snow density, temperature and structure. Fresh snow density lies between $0.075 \div 0.20$ g/cm^3, while seasoned and compacted snow density is between $0.30 \div 0.40$ g/cm^3; the transformation of fresh to seasoned snow is a one-way process.

Soil deformation because of vertical load

The application of a vertical load onto soils creates deformation; we will consider the application of uniform loads, as we could ideally have with an infinitely flexible tire (membrane model). Let us consider Fig. 2.7, which represents in a qualitative way what could happen by pressing a pad, applying a pressure p that is defined as the ratio between the applied load and the pad surface.

This pad simulates the effect of the contact patch of a tire, pressed against the soil by the weight of the vehicle.

There is no difference in this description related to the different types of soils we have classified.

With relatively small pressure, the soil is cut along the perimeter of the pad and is tamped down; the pressure reaction of the soil is concentrated on the edge of the cut and increases as soil cohesion increases. A nucleus of tamped down soil builds up below the pad and moves in the direction of the applied load, loading lower layers of soil. In a first phase deformation is almost proportional to applied pressure; soil reaction pressure is shown qualitatively below the diagram, keyed to the corresponding part of the $h(p)$ diagram.

As pressure is increased more significant volumes of soil reach plasticity, as is shown in the central image for phase 2. When the entire volume of soil displaced by the pad has reached the plastic condition it reacts with the same pressure at any value of sinking. The position and extension of phase 2 depends on initial soil density.

Figure 2.8 shows some diagrams of sinking h as a function of applied pressure p, for various kinds of soil, characterized by the qualitative behavior we have described.

www.TechnicalBooksPDF.com

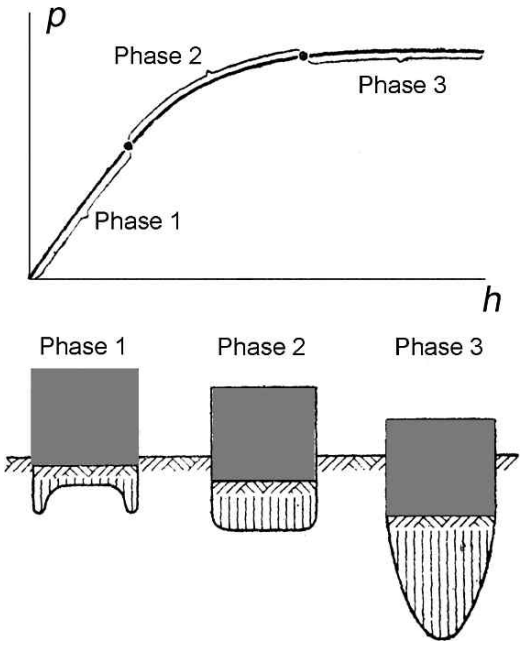


FIGURE 2.7. Main phases of deformation of an ideal homogeneous soil, expressed by sinking h as a function of applied pressure p. The lower diagrams show qualitatively the distribution of stress in the soil.

Diagram A refers to a cohesive soil (a: humid soil, b: plastic soil); diagram B refers to a sandy soil (a: loose sand, about 200 mm thick; b: a compacted layer); diagram C refers to a layer of peat (a: pressing pad of 4 m^2 of surface; b: pressing pad of about 0.4 m^2); diagram D refers to a snowy ground (a: fresh snow, density of 0.15 g/cm^2, b: compacted snow, density of 0.20 g/cm^2).

Diagram C shows haw the surface of the pad can influence the value of sinking at the same pressure. In particular the larger the surface, the greater the sinking; this happens because larger surfaces increase the amount of soil interested by tamping.

Please take note of the different order of magnitude for the pressure of cohesive soils, which demonstrates their superior mechanical properties; considering a tire with a patch pressure of about 0.2 MPa, we can immediately draw first approximation conclusions about the possibility of driving a vehicle on different kinds of soil.

In the case of fresh snow, we show in Fig. 2.9 several sinking curves for different average dimension d of the pad ($d = \sqrt[2]{A}$, where A is the pad area). Curve a corresponds to a pressure $p = 0.1$ MPa; curve b to $p = 0.08$ MPa, curve c to $p = 0.06$ MPa and curve d to $p = 0.04$ MPa.

Similar curves can be found for different kinds of soil, even if at times there is a minimum point for low values of d.

www.TechnicalBooksPDF.com

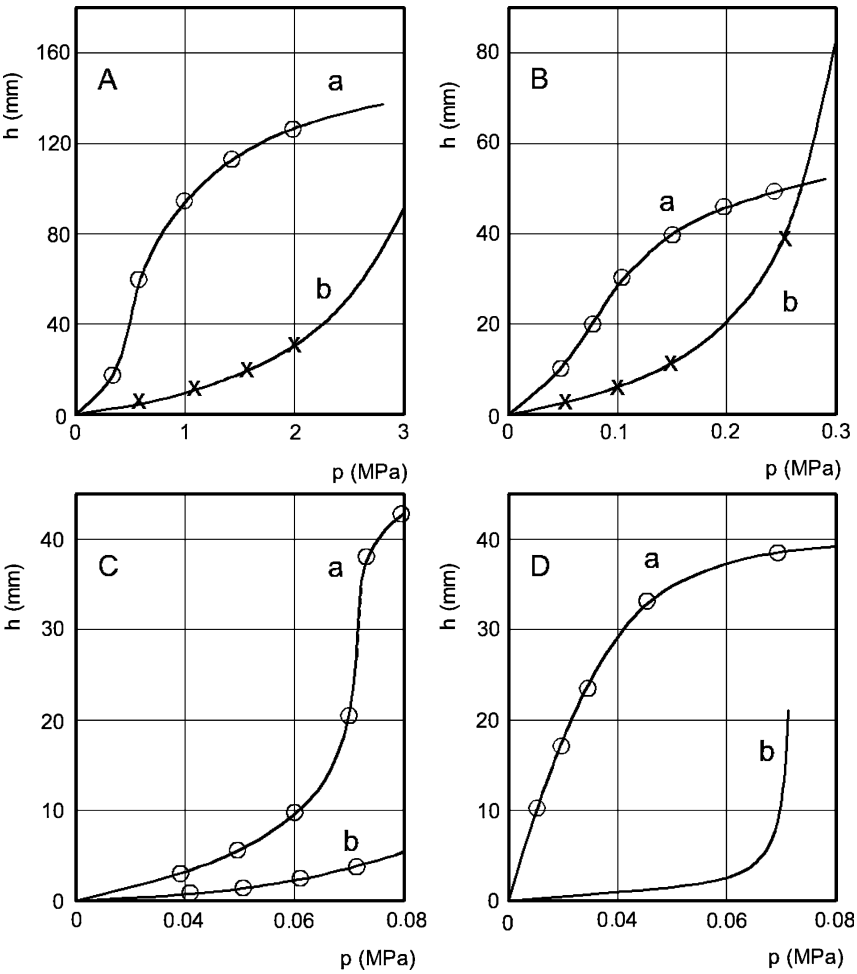


FIGURE 2.8. Diagram of sinking h as a function of applied pressure p for various kind of soil. Diagram A refers to a cohesive soil (a: humid soil, b: plastic soil); diagram B refers to a sandy soil (a: loose sand, about 200 mm thick; b: a compacted layer); diagram C refers to a layer of peat (a: pressing piston of 4 m^2 of surface; b: pressing piston of about 0.4 m^2); diagram D refers to a snowy ground (a: fresh snow, density of 0.15 g/cm^2, b: compacted snow, density of 0.20 g/cm^2).

In the first part of the curve in Fig. 2.7a (phase 1), where the soil behaves as an elastic material, it is possible to define a coefficient of proportionality, sometimes called the stiffness or *bedding number* of the soil

$$k = \frac{p}{h} \ . \tag{2.3}$$

www.TechnicalBooksPDF.com

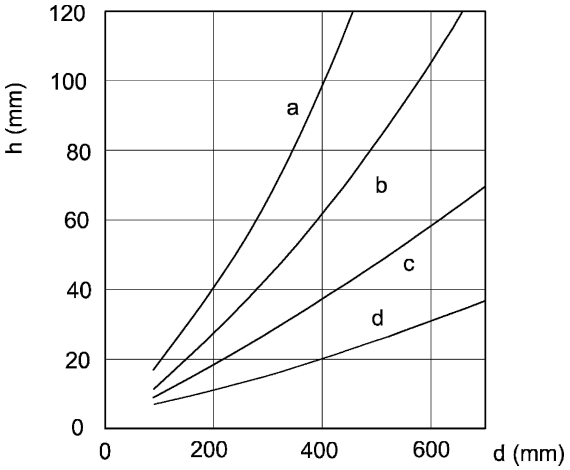


FIGURE 2.9. Effect of the average piston dimension d on sinking h, for different values of pressure p, on snow (curve a: $p = 0.1$ MPa; curve b: $p = 0.08$ MPa; curve c: $p = 0.06$ MPa; curve d: $p = 0.04$ MPa).

The stiffness k does not depend solely on the elastic characteristics of the ground (its Young's modulus E and its Poisson's ratio ν), but also on the area and the shape of the loading pad.

Consider a pad exerting a constant pressure p on the soil. In Phase 1 the sinking h reduces to the elastic sinking h_e defined as

$$h_0 = \frac{p}{k} \ . \tag{2.4}$$

The sinking h in the subsequent phases 2 and 3 may be expressed as

$$h = \frac{h_e p_0}{p_0 - p} = h_e \frac{1}{1 - p/p_0} \ , \tag{2.5}$$

where h and h_0 have the meaning of effective deformation and the contribution of elasticity, to total deformation interpreted by the previous formula; p_0, *plasticity pressure*, is the value of pressure over which plasticity takes place, as in phase 3. Obviously this formula is valid for $p \geqslant p_0$ only.

Soil deformation because of shear forces

The simultaneous presence of side forces applied to the pad causes an increase of sinking at the same pressure; some experimental behaviors are reported in Fig. 2.10.

The diagrams show pad sinking h due to the contemporary presence of normal pressure p and shear pressure τ, for a sandy soil (Figure A) and a clayey soil (Figure B). Curves a, b and c correspond to pressure p of 0.02, 0.03 and 0.04 MPa.

www.TechnicalBooksPDF.com

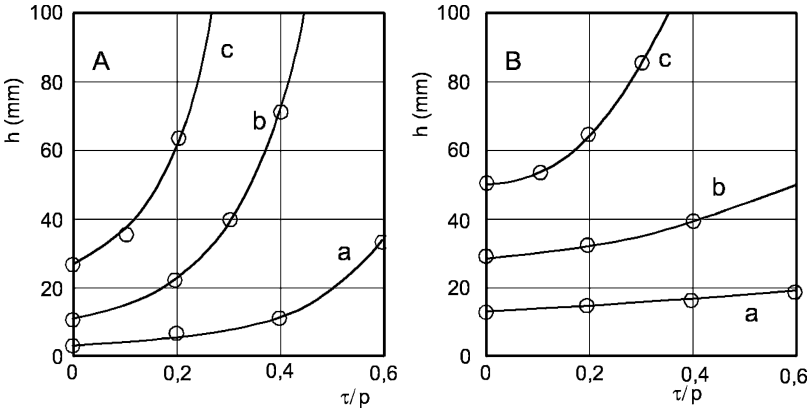


FIGURE 2.10. Piston sinking h due to the simultaneous presence of normal pressure p and shear pressure τ, for a sandy soil (A) and a clayey soil (B). Curves a, b and c correspond to pressure p of 0.02, 0.03 and 0.04 MPa.

The tangential force the pad can exert on the ground is limited by cohesive and friction phenomena occurring at the soil-pad interface. If the pad does not sink into the ground, the maximum tangential force is

$$F_{t_{\max}} = AC_0 + F_n \tan \Phi_0 \ , \tag{2.6}$$

where C_0 is the contribution of cohesion, or *cohesivity* and Φ_0 is the *friction angle* between particles; it can be practically measured by evaluating the semi-aperture of the cone, which approximates the shape of a pile of this soil in the loose state.

The shear resistance of the considered soil τ_o is then

$$\tau_0 = \frac{F_{t_{\max}}}{A} = C_0 + p \tan \Phi_0 \ . \tag{2.7}$$

If the pad sinks into the ground, a further component of the force, the *bulldozing* force, must be added to those caused to cohesion and friction.

Mechanical properties of different soils

Figure 2.11 shows the main parameters of identification of the mechanical behavior of a natural ground:

- Elastic modulus E
- Plasticity pressure p_0
- Friction angle Φ_0
- Cohesivity C_0

www.TechnicalBooksPDF.com

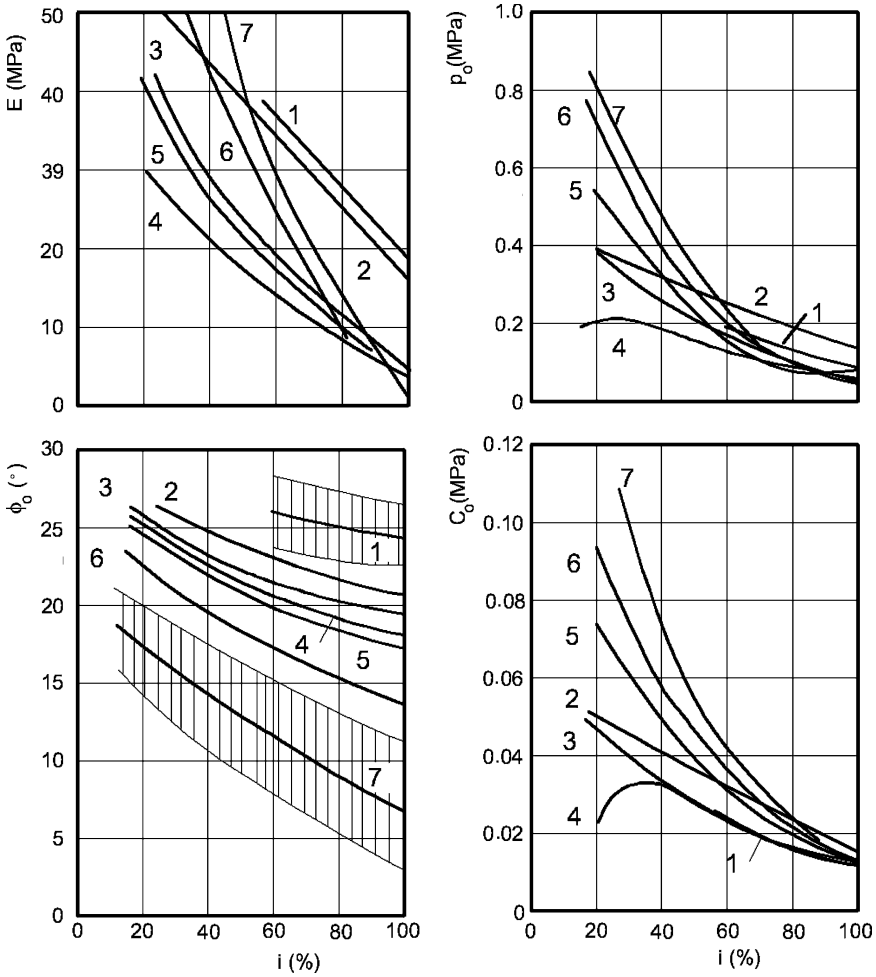


FIGURE 2.11. Primary identification parameters of the behavior of a natural ground, as a function of the relative content of water i; curves with number 1 refer to a sandy loose soil, with 7 to a clayey soil, while numbers 2 to 6 refer to sandy soils with increasing sand content.

to calculate sinking h as a function of normal pressure p and shear pressure τ. In addition, τ_0 allows to predict the maximum longitudinal force (driving or braking) that the ground can withstand.

These parameters are a function of the relative content of water i; curves with number 1 refer to a sandy loose soil, with 7 to a clayey soil, while numbers 2 to 6 refer to sandy soils with increasing sand content.

The deterioration of these parameters for increasing water content is noticeable; as a reference a ploughed and hoed soil in Northern Europe can contain up to 100% water at the end of ice melt and 30% water in the summer.

www.TechnicalBooksPDF.com

TABLE 2.2. Primary mechanical properties of snow, as a function of its state and external temperature.

T(°C)	Physical state	Fresh	Compacted by wind	Compressed
–	Density (g/cm^3)	0.2	0.4	0.6
−5	Cohesion (MPa)	0.05	0.07	0.19
−10	Cohesion (MPa)	0.06	0.09	0.22
−5	Friction angle (deg)	18	22	31
−10	Friction angle (deg)	19	23	26
−5	Elastic modulus (MPa)	$0.4 \div 0.6$	$1.5 \div 2.0$	∼7.5
−10	Elastic modulus (MPa)	$0.6 \div 0.8$	$2.5 \div 3.0$	∼10

The higher cohesion of clayey soils and the higher friction angle of sandy soils can be noticed.

Table 2.2 illustrates the primary mechanical properties of snow, as a function of its state and external temperature.

What we have reported allows the following problems to be solved:

- Determining whether a particular ground can withstand vehicle wheel pressure, without interfering with the chassis
- Predicting the sinking of the contact patch
- Defining the maximum longitudinal or side force, by calculation of the shear resistance τ_0

The contributions to tire patch displacement (vertical deformation, longitudinal slip and sideslip angle) will be described in the following paragraphs; such contributions are relevant on paved roads, but can be neglected on natural grounds.

Rolling resistance must also be considered. A tire rolling on unpaved roads or natural ground is subject to the following forces.

- Resistance due to elastic hysteresis of the tire; this force, which we will describe in the following paragraphs, is caused by the energy dissipated by the deformation of the contact patch.
- Resistance due to elastic hysteresis of soil; this force is caused by the mechanical work of plastic deformation, as we have described.
- *Bulldozing resistance* due to the fact that a certain amount of soil is pushed in front of the contact patch of the tire.

The resistance due to tire hysteresis is slightly different than that measured on paved grounds because the patch flattening is lower; other contributions could be calculated beginning with the physical properties of the soil in question: These can increase rolling resistance of an order of magnitude.

www.TechnicalBooksPDF.com

TABLE 2.3. Values of f_0 for different kinds of natural grounds.

Type of road	f_0
Flat snowy road	0.025–0.040
Natural flat ground	0.08–0.16
Flat sandy ground	0.15–0.30

Table 2.3 shows some values of rolling resistance f_0[3] for most common types of natural ground. We invite the reader to compare these values with those reported in Table 2.4 for paved roads.

The following chapters will refer to the tire on paved roads only.

2.3 ROLLING RADIUS

Consider a wheel rolling on a level road with no braking or tractive moment applied to it and its mean plane perpendicular to the road. While the relationship between the angular velocity Ω and the forward speed V of a rolling rigid wheel of radius R is simply

$$V = \Omega R,$$

for a pneumatic tire an effective rolling radius R_e can be defined as the ratio between V and Ω:

$$R_e = V/\Omega \ . \tag{2.8}$$

The effective *rolling radius* is defined as the radius of a rigid wheel that travels and rotates at the same speed as the pneumatic wheel.

The wheel-road contact is far from being a point-contact. The tread band is also compliant in a circumferential direction; as a consequence radius R_e coincides neither with the loaded radius R_l nor with its unloaded radius R. The centre of instantaneous rotation is therefore not coincident with the centre of contact A (Fig. 2.12).

Because of longitudinal deformations of the tread band, the peripheral velocity of any point of the tread varies periodically. When it nears the point at which it enters the contact zone it slows and a circumferential compression consequently results. In the contact zone sliding between tire and road is quite limited.

The peripheral velocity of the tread (relative to the wheel centre) in that zone coincides with the velocity of the centre of the wheel V. After leaving the contact zone, the tread regains its initial length and its peripheral velocity ΩR is restored.

As a consequence of this mechanism, the spin speed of a wheel with a pneumatic tire is smaller than that of a rigid wheel with the same loaded radius R_l and travelling at the same speed:

[3]Rolling resistance coefficient f_0 will be defined later.

www.TechnicalBooksPDF.com

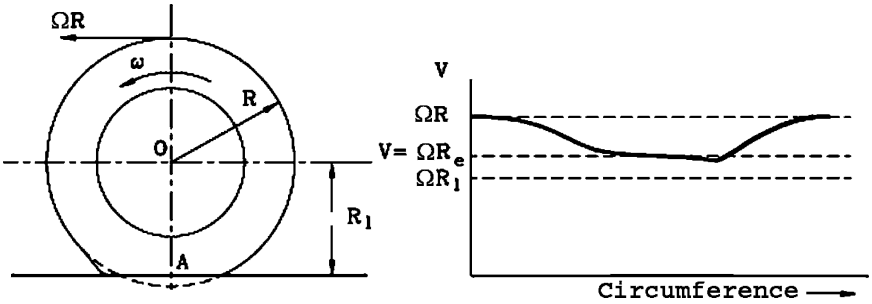


FIGURE 2.12. Tire rolling on a flat road; geometric configuration and peripheral speed at the contact area.

$$R_l < R_e < R\,.$$

The centre of rotation of the wheel then lies a short distance under the surface of the road.

Owing to their lower vertical stiffness, radial tires have a lower loaded radius R_l than bias-ply tires with equal radius R, but their effective rolling radius R_e is closer to the unloaded radius, as the tread is circumferentially stiffer.

For instance in a bias-ply tire R_e can be about 96% of R while R_l is 94% of it; in a radial tire R_e and R_l can be, respectively, 98% and 92% of R.

The effective rolling radius depends on many factors, some of which are determined by the tire, such as the type of structure, the wear of the tread, and the working condition, which can involve inflation pressure, load, speed and other factors.

An increase of the vertical load F_z and a decrease of the inflation pressure p lead to similar results: A decrease of both R_l and R_e. With increasing speed, the tire expands under centrifugal forces, and consequently R, R_l and R_e increase. This effect is larger in bias-ply tires while, owing to the greater stiffness of the tread band, radial tires expand to a limited, and usually negligible extent. As will be shown in the following sections, any tractive or braking torque applied to the wheel will cause strong variations of the effective rolling radius.

The diagram of rolling radius as a function of speed for tires 155 D 15 with cross ply and 155 R 15 with radial ply is shown in Fig. 2.13. Diagrams are obtained by having the tire roll inside a 3.8 m diameter steel drum. As we will show later any, variation in tractive or braking force affects rolling radius.

2.4 ROLLING RESISTANCE

Consider a wheel rolling freely on a flat surface. If both the wheel and the road were perfectly undeformable, there would be no resistance and consequently no need to exert a tractive force. In the real world, however, perfectly rigid bodies do not exist and both road and wheel are subject to deformation in the contact zone.

www.TechnicalBooksPDF.com

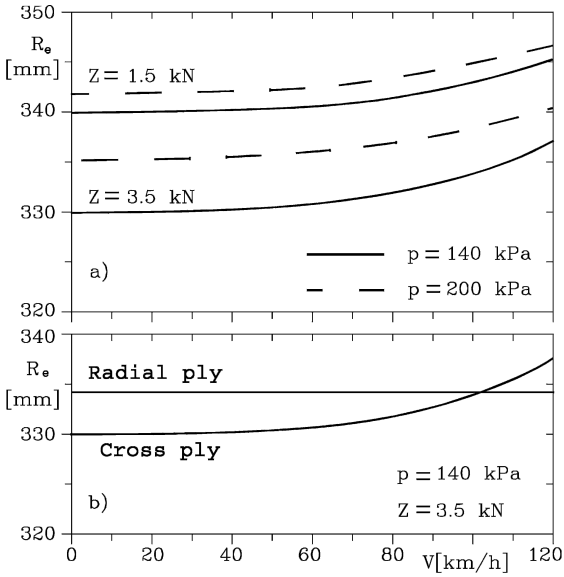


FIGURE 2.13. Rolling ratio R_e as a function of V for a cross ply tire 155 D 15 (a) and for a radial ply tire 155 R 15 (b).

During tire motion new material enters continuously into this zone and is deformed, springing back to its initial shape when leaving the zone. To produce this deformation some energy must be expended that is not completely recovered at the end of the contact because of the internal damping of the material.

This energy dissipation is what causes rolling resistance. It is thus clear that it increases with increasing deformation and, primarily, with decreasing elastic return. A steel wheel on a steel rail has a lower rolling resistance than a pneumatic wheel, while the motion on compliant soil causes greater resistance than that on a rigid surface. From this viewpoint, a wheel on compliant soil is always in the position of a wheel attempting to climb out of the pit it is digging itself into (Fig. 2.14a).

In the case of pneumatic tires rolling on tarmac or concrete, the deformations are almost always localized only in the wheel so that the energy dissipated in the tires governs the phenomenon. Other mechanisms, such as slight sliding between road and wheel, aerodynamic drag on the disc and friction in the hub are responsible for a small contribution to overall resistance, on the order of a few percent.

The distribution of the contact pressure, which at standstill was symmetrical with respect to the centre of the contact zone, becomes unsymmetrical when the wheel is rolling and the resultant F_z moves forward (Fig. 2.14b), producing a torque $M_y = -F_z \Delta x$ with respect to the rotation axis. Rolling resistance is caused by this torque, along with the small contributions of the resistance in the hub and aerodynamic drag.

www.TechnicalBooksPDF.com

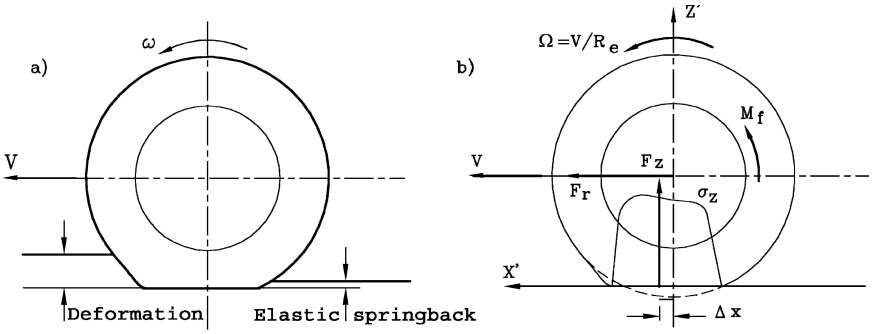


FIGURE 2.14. (a) Rolling tire on a deformable surface: ground deformation and spring back. (b) Forces F_z ed F_r and contact pressure σ_z in a rolling tire.

The two mentioned ways of viewing rolling drag are equivalent because resultant F_z is displaced forward to the centre of contact owing to the energy dissipations occurring in the deformed parts of the wheel and possibly the ground.

Rolling resistance is defined by the above mentioned SAE document J670 as the force which must be applied to the wheel at the wheel centre with a line of action parallel to the X' axis so that its moment with respect to a line through the centre of tire contact and parallel to the spin axis of the wheel will balance the moment of the tire contact forces about this line.

Here this definition is used, with two small modifications.

First, the force is changed in sign, so that a true resisting force, i.e. a force acting in a direction opposite to the speed, is obtained.

Second, the aerodynamic drag moment on the wheel and the resisting torque applied to the hub are included, in order to factor these two effects into the overall rolling resistance.

To keep a free wheel spinning, a force at the wheel-ground contact is required. Some of the available traction is used: On the idle wheel, to supply a torque which counteracts the total moment M_y, and on the driving wheels, which must supply a tractive force against the rolling resistance of the former.

The driving torque is directly applied to the driving wheels through the driveshafts to overcome rolling resistance moment. The rolling resistance of the driving wheels thus does not involve forces acting at the road-wheel contact and does not use any of the available traction.

This is particularly important in motion on compliant ground, which is usually characterized by high rolling drag and low available traction: If all wheels are driving wheels, rolling drag can be overcome directly by the driving torque; if some of the wheels are idle, the traction the driving wheels can supply may not be sufficient to overcome the drag of the idle wheels and motion may be impossible, even on level road.

Consider a free rolling wheel on level road with its mean plane coinciding with the $X'Z'$-plane, i.e. with $\alpha = 0$, $\gamma = 0$ (Fig. 2.14). Assuming that no traction or braking moment other than moment M_f due to aerodynamic drag and bearing

www.TechnicalBooksPDF.com

drag is applied to the wheel, the equilibrium equation in steady state rolling, solved in the rolling resistance F_r, is:

$$F_r = \frac{-F_z \Delta x + M_f}{R_l} \ . \tag{2.9}$$

It must be noted that both rolling resistance F_r and drag moment M_f are negative.

In the case of driving wheels, moment M_f must be substituted by the difference:

$$M_m - |M_f| \ ,$$

between the driving and resistant moments.

If this difference is positive and greater than moment $F_z \Delta x$, force F_r is positive, i.e. the wheel is exerting a driving action.

Equation (2.9) is of limited practical use, as Δx and M_f are not easily determined. For practical purposes, rolling resistance is usually expressed as:

$$F_r = -f F_z \ , \tag{2.10}$$

where the rolling resistance coefficient f must be determined experimentally.

The minus sign in Eq. (2.10) derives from the fact that the rolling resistance coefficient is traditionally expressed by a positive number.

Coefficient f depends on many parameters, such as the travelling speed V, the inflation pressure p, the normal force F_z, the size of the tire and of the contact zone, the structure and the material of the tire, the working temperature, the road conditions and, last but not least, the forces F_x and F_y exerted by the wheel.

2.4.1 Effect of speed

The rolling resistance coefficient f generally increases with the speed V of the vehicle, slowly at the beginning and then at an increased rate (Fig. 2.15).

The law $f(V)$ can be approximated by a polynomial expression of the type:

$$f = \sum_{i=0}^{n} f_i V^i \ . \tag{2.11}$$

Generally speaking, two terms of Eq. (2.11) are considered sufficient for approximating experimental data in a satisfactory way, at least up to the speed at which f starts to grow at a very high rate (Fig. 2.15). The following expressions can be used:

$$f = f_0 + kV \tag{2.12}$$

or:

$$f = f_0 + KV^2 \ . \tag{2.13}$$

www.TechnicalBooksPDF.com

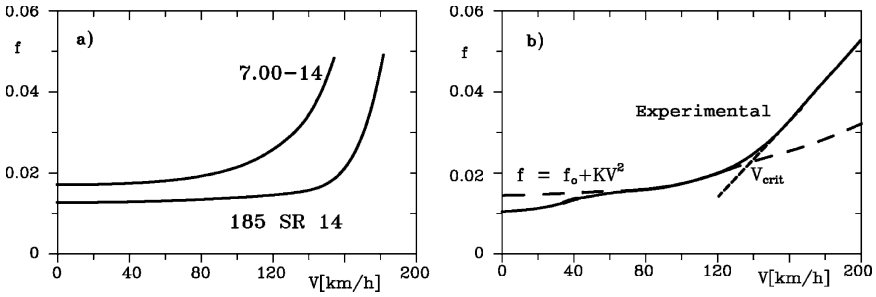


FIGURE 2.15. Diagram of rolling resistance f as a function of speed V. (a) Measured on a cross ply tire 175 14 and on a radial tire 185 R 14; (b) experimental curve (radial tire 135 R14, inflated at 190 kPa, with vertical load of 340 kN) compared with Eq. (2.13).

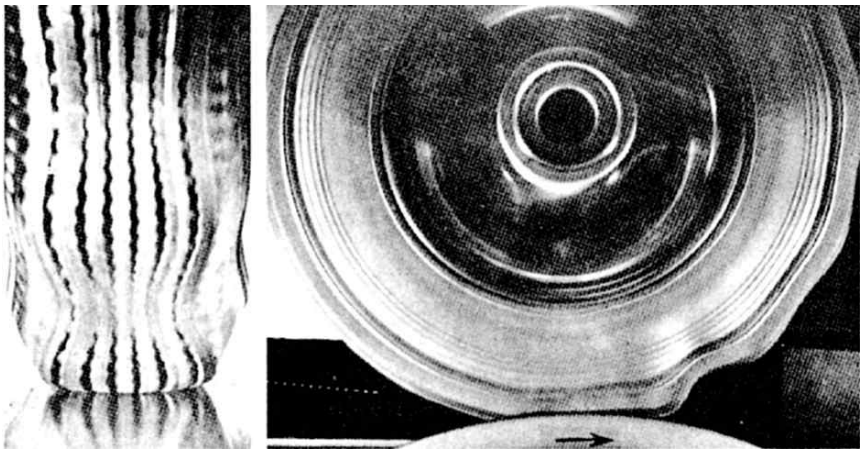


FIGURE 2.16. Stationary waves on a rolling tire over its critical speed.

The second is generally preferred; it will be used throughout this book. The values of f_0 and K must be measured for a particular tire; as an example the tire of Fig. 2.15b is characterized, in the test conditions reported, by the values: $f_0 = 0.013$, $K = 6.5 \times 10^{-6}$ s^2/m^2.

On more recent tires rubber mix and fillers have been developed that are able to reduce rolling resistance; values of $f_0 = 0.008$ can be achieved.

The speed at which the curve $f(V)$ shows a sharp bend upward is generally said to be the *critical speed* of the tire. Its presence can be easily explained by vibratory phenomena that take place in the tire at high speed, such as those clearly visible in the pictures of a tire rolling at high speed against a drum shown in Fig. 2.16. The standing waves that propagate along the circumference of the tire from the contact zone are clearly visible. The tread band vibrates both in its plane and in the direction of the axis of the wheel.

www.TechnicalBooksPDF.com

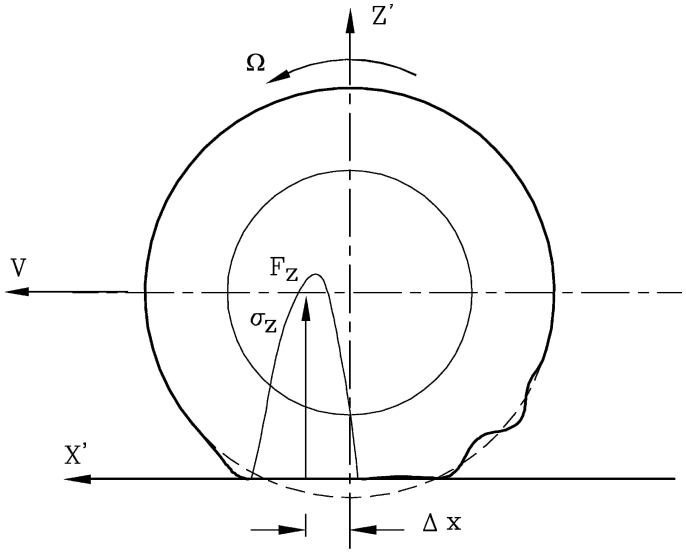


FIGURE 2.17. High speed stationary waves (A. Morelli, *Costruzioni automobilistiche*, in *Enciclopedia dell'ingegneria*, ISEDI, Milano, 1972).

The increase in rolling resistance that is linked with the occurrence of standing waves is easily explained by the fact that their wavelength is not greatly different from the length of the contact zone (Fig. 2.17). In the trailing part of the contact zone the tread thus has a tendency to lift from the ground, or at least to decrease its pressure on it. The pressures concentrate consequently in the leading zone of the contact and their resultant moves forward, with an increase of the moment $F_z \Delta x$.

The critical speed of the tire, i.e. the speed at which such vibrations become important, must be considered as the speed at which the tire stops functioning normally. As a result it should never be exceeded or even approached in the normal use of the vehicle. Above this speed strong overheating takes place; as most of the increase of the rolling power is converted into heat, the increase in temperature can quickly cause the destruction of the tire.

The critical speed is influenced by many parameters and it is one of the factors that must be taken into account in the choice of the tires for a particular vehicle.

2.4.2 Effect of material nature and structure

The type of structure and the material used in the construction of the tire play an important role in determining the rolling resistance and critical speed. Generally speaking, radial tires show a value of f about 20% lower than bias-ply types (Fig. 2.15) and a higher value of the critical speed.

www.TechnicalBooksPDF.com

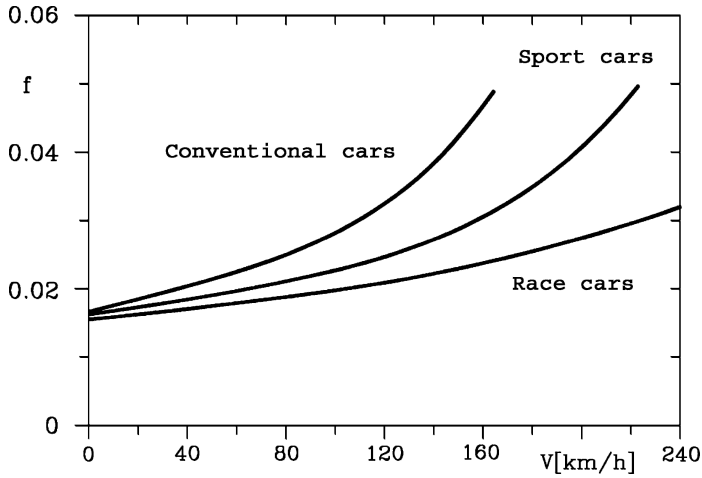


FIGURE 2.18. Rolling resistance coefficient f and speed V for different purpose tires.

Even between tires of the same type, significant differences in rolling resistance may be observed, as it is possible to optimize the structure (number of plies, their orientation etc.) in order to obtain suitable characteristics for any particular application.

In Fig. 2.18 some typical curves are reported, for example, referring to conventional cars, sport cars and race cars. It can be seen that critical speeds are quite different in the three cases.

Notable differences can be seen between car and industrial vehicle tires. The latter present values of f_0 that are much lower, down to $0.005 \div 0.008$, and show a limited increase of f with speed ($K \approx 0$), or sometimes a slight reduction (Fig. 2.19).

The nature of the material used has great importance, as different rubber compositions are characterized by different values of internal damping and by different dependence of the latter upon the loading frequency. Natural rubbers generally show low damping when compared with synthetic rubbers. This results in lower rolling resistance but also in lower critical speed.

The type and quantity of the other chemical components added to rubber also have an important influence on damping and consequently on rolling resistance. In Fig. 2.20 two curves are shown for same size tires, one made with natural rubber, the other by synthetic.

2.4.3 Effect of tread wear

In Fig. 2.21 curves $f(V)$ are shown for two tires, cross ply and conventional, in new conditions. To simulate tread wear part of the tread has been artificially removed and the tests repeated.

www.TechnicalBooksPDF.com

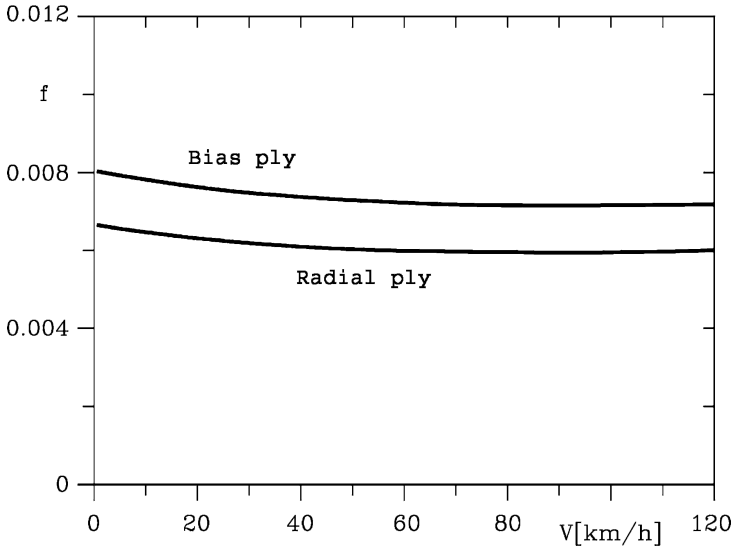


FIGURE 2.19. Diagram of f with V on industrial vehicle tires 300 20, pressure $p = 750$ kPa.

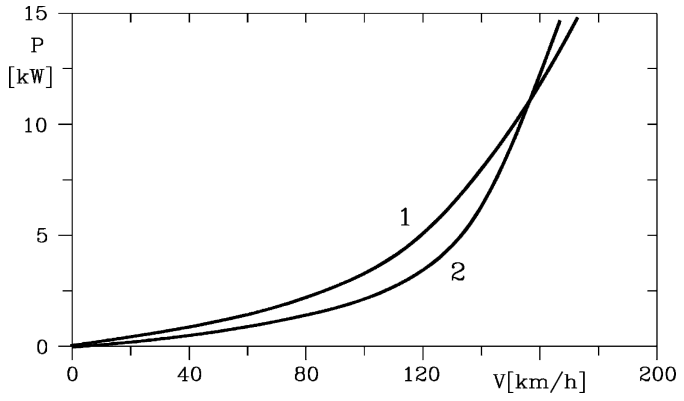


FIGURE 2.20. Rolling power as a function of speed. Tires 135 l4, $p = 175$ kPa, $F_z = 3{,}7$ kN. Curve 1: synthetic rubber; curve 2: natural rubber.

In the case of bias-ply tires the rolling resistance decreases with wear, and its behavior at high speed improves. This behavior can be ascribed to the fact that deformations are localized in a small zone surrounding the contact zone and consequently hysteresis losses take place mainly in the tread band. Also vibratory phenomena involve primarily the zone immediately adjacent to the tread band. A decrease of the vibrating mass has the consequence of increasing the natural frequency and hence the critical speed.

In radial tires rolling resistance decreases with wear, but behavior at high speed is degraded. In radial tires deformations are more evenly distributed

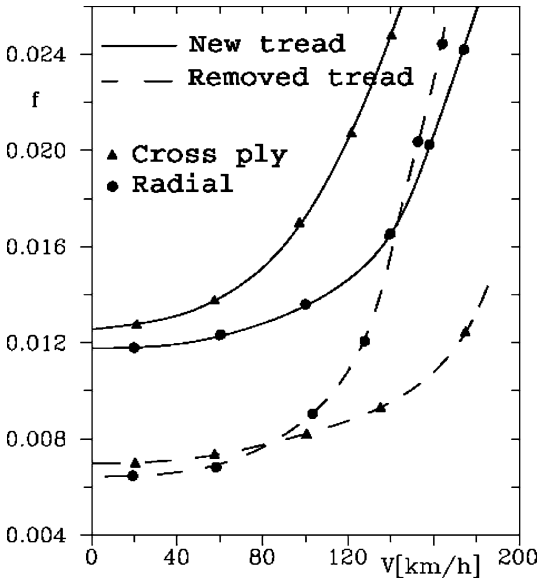


FIGURE 2.21. Effect of wear on f.

throughout the structure, as the stiffness of the sidewalls is low; with a decrease of the mass of the tread band, the centrifugal stiffening of the whole structure decreases, and vibratory phenomena become more important.

2.4.4 Effect of operating temperature

Internal damping of rubber decreases with increasing temperature and consequently rolling resistance, which is mainly due to hysteresis losses, decreases. That small part of rolling resistance that is due to localized sliding in the contact zone decreases as the friction coefficient decreases.

The decrease of rolling resistance tends to stabilize the temperature of the tire as an increase of the temperature causes a decrease of power dissipation and consequently the rate at which heat is generated within the tire.

Some curves $f(V)$ obtained at constant temperature are shown in Fig. 2.22a, while in Fig. 2.22b the rolling coefficient of the same tire maintained at each speed at the equilibrium temperature is shown. The curve is compared with that obtained by maintaining the tire at a temperature corresponding to low-speed running, as occurs at start-up of the vehicle, before equilibrium temperature is reached.

The decrease in time of the rolling resistance and the increase of the temperature in a tire rolling at 185 km/h is shown in Fig. 2.22c, while in Fig. 2.22d the equilibrium temperature is plotted as a function of the speed. The last two plots have been obtained on a Nylon 185 13 radial tire, with $F_z = 4$ kN and $p = 150$ kPa, rolling against a roller of 2.5 m diameter.

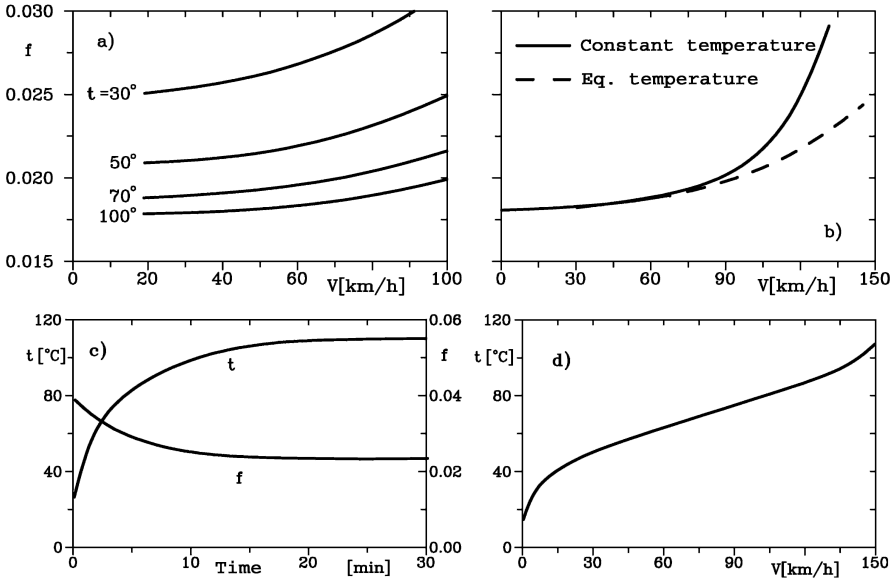


FIGURE 2.22. (a) Rolling resistance coefficient at constant temperature. (b) Comparison between $f(V)$ at constant temperature and at equilibrium temperature for each speed. (c) Reduction of rolling resistance coefficient and temperature increase for increasing time at a speed of 185 km/h. (d) Equilibrium temperature as a function of speed.

The temperature was measured using a thermocouple inset in the tire body and reflects the temperature of the material, which is higher than that of the air. As the tests were performed on a roller, the increase of temperature and resistance are somewhat higher than those that would be obtained on the road.

2.4.5 Effect of inflation pressure and vertical load

Generally speaking an increase of inflation pressure or a reduction of the normal force F_z acting on the wheel causes the rolling resistance to decrease and the critical speed to increase. This effect also tends to stabilize the temperature, as an increase of the latter causes an increase of the pressure which in turn decreases power dissipation and heat generation.

In Fig. 2.23a the $f(V)$ curves are shown, for different values of inflation pressure p. The behavior of the tire at different vertical loads F_z and at different pressures p is shown in Fig. 2.23b.

In order to take into account the influence of both load and pressure on the rolling resistance coefficient, the following empirical formula suggested by the SAE can be used:

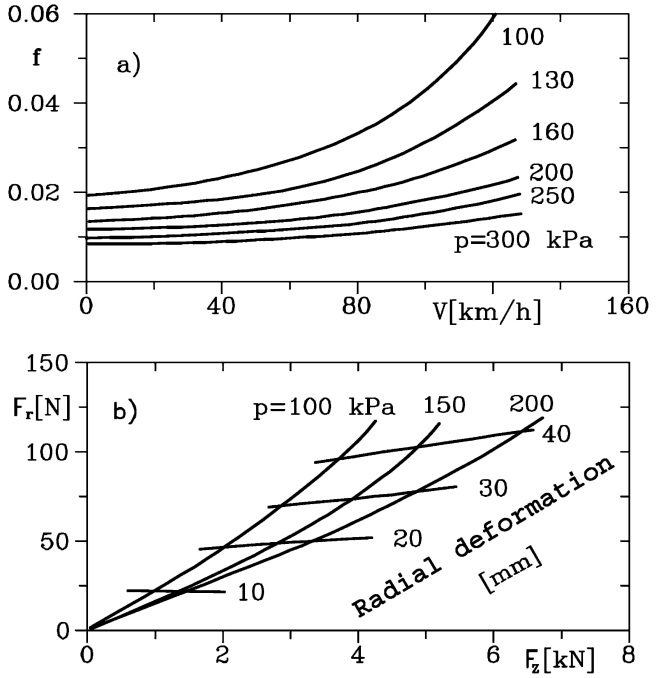


FIGURE 2.23. (a) Effect of the inflation pressure on $f(V)$; (b) behavior of a 165 13 tire at different values of F_z and inflation pressure p. Speed $V = 30$ km/h.

$$f = \frac{K'}{1000}\left(5,1 + \frac{5,5 \times 10^5 + 90F_z}{p} + \frac{1100 + 0,0388F_z}{p}V^2\right) , \qquad (2.14)$$

where coefficient K' takes the value 1 for conventional tires and 0.8 for radial tires. The normal force F_z, the pressure p and the speed V must be expressed respectively in N, N/m^2 (Pa) and m/s. The dependence of coefficient f on the speed V is the same as expressed by Eq. (2.13).

It must, however, be noted that for each tire the inflation pressure p is determined as a function of the force F_z by design considerations, and that it is impossible to increase the pressure in order to reduce the rolling resistance.

2.4.6 Effect of tire size

The two geometrical parameters which have more influence on the rolling resistance are the radius of the tire and the aspect ratio H/W. An increase of the former and a decrease of the latter cause a decrease of rolling resistance and an increase of the critical speed.

The decrease of the aspect ratio is favorable as it causes an increase of the stiffness of the sidewalls and decreases the deformation under load, which in turn

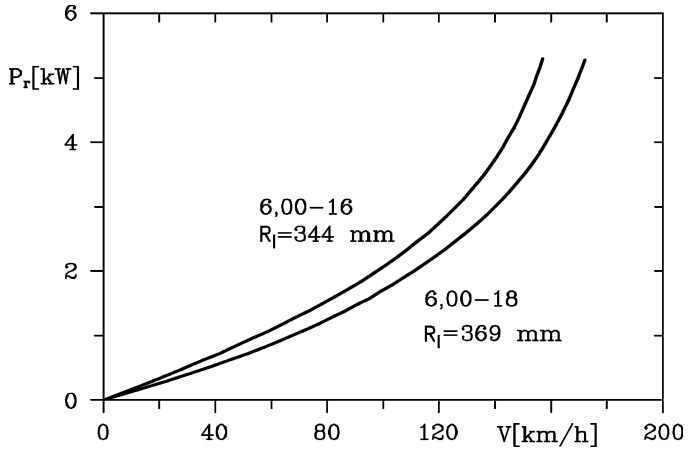


FIGURE 2.24. Influence of rolling radius on rolling resistance power P_r. $F_z = 4.75$ kN, $p = 250$ kPa.

lowers the hysteresis losses. This ratio can be as low as 0.4 on radial tires used on fast modern cars, while values of $0.7 \div 0.8$ were most common in the past.

Figure 2.24 shows the rolling power dissipated by two tires of different size.

2.4.7 Effect of road

As a first approximation, the nature and conditions of the road are taken into account by choosing an appropriate value of f_0, i.e. by moving the curve $f(V)$ in the direction of the f axis. Some values of f_0 for different road types are shown in Table 2.4.

2.4.8 Effect of wheel sideslip angle

If the tire travels with a sideslip angle α, as is the case any time it exerts a side force F_y or as a consequence of toe angle, a strong increase of rolling resistance can be expected. The force in the mean plane of the wheel increases but above all the transversal force F_y has a component that adds to the rolling resistance (Fig. 2.25). The rolling resistance is by definition the component of the force due to the road-tire contact directed as the velocity V; it can thus be expressed as:

$$F_r = F_x \cos(\alpha) + F_y \sin(\alpha) \; . \tag{2.15}$$

If the component in the plane of symmetry of the wheel F_x were independent of the sideslip angle and the cornering force F_y were linear with it ($F_y = -C\alpha$), for small values of α the rolling resistance would follow a quadratic law:

$$|F_r| = |F_x| + C\alpha^2 \; . \tag{2.16}$$

TABLE 2.4. Values of f_0 for different kinds of roads.

Type and conditions of road	f_0
Very good concrete	0.008–0.010
Very good tarmac	0.010–0.0125
Average concrete	0.010–0.015
Very good paved road	0.015
Very good macadam	0.013–0.016
Average tarmac	0.018
Bad concrete	0.020
Good paved road	0.020
Average macadam	0.018–0.023
Bad tarmac	0.023
Dusty macadam	0.023–0.028
Good stone paved road	0.033–0.055
Very good dirt road	0.045
Bad stone paved road	0.085

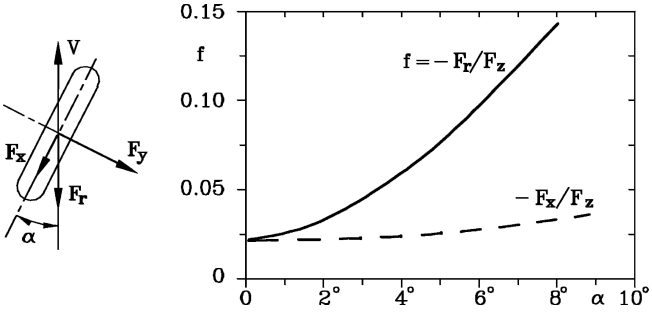


FIGURE 2.25. Rolling resistance coefficient as function of the wheel side slip angle α. Tire 185 14, $F_z = 4$ kN, $p = 170$ kPa.

2.4.9 Effect of camber angle

If the mean plane of the wheel is not perpendicular to the ground a component of the aligning torque M_z (see Sec. 2.7) contributes to rolling resistance. Equation (2.9) becomes:

$$F_r = \frac{-F_z \Delta x \cos(\gamma) - M_z \sin(\gamma) + M_f}{R_l} \,. \tag{2.17}$$

This effect is usually very small, due to the fact that γ is usually small. It is, however, dependent on the sideslip angle α through the aligning torque M_z.

Effect of longitudinal force

Rolling resistance can also be defined when tractive or braking moments are applied to the wheel. In this case the power dissipated by rolling resistance $F_r V$ can be expressed as:

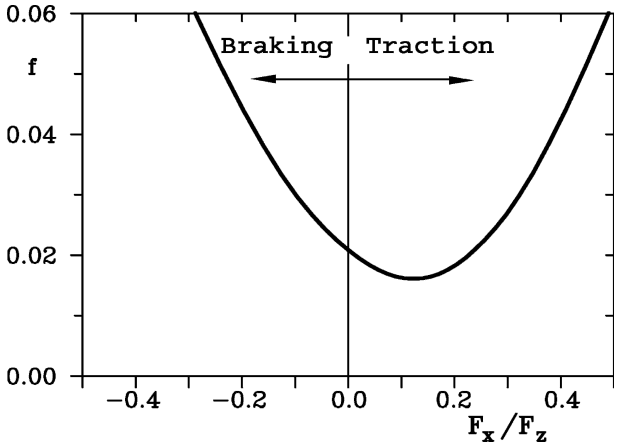


FIGURE 2.26. Rolling resistance coefficient as function of braking and traction force.

$$|F_r|V = \begin{cases} |F_f|V - |M_f|\Omega & \text{(braking)} \\ M_t|\Omega - |F_t|V & \text{(traction)} \end{cases} \tag{2.18}$$

where F_f, F_t, M_f and M_t are, respectively, the braking and tractive forces and moments. Equation (2.18) should be applied only in constant speed motion, because they do not include the tractive (braking) moments needed to accelerate (decelerate) rotating parts.

The trend of the rolling resistance coefficient as a function of the longitudinal force F_x is shown in Fig. 2.26. The increase of rolling resistance is not negligible in the case of strong longitudinal forces, particularly braking. This is due primarily to the fact that the generation of longitudinal forces is always accompanied by the presence of sliding in at least part of the contact zone.

The minimum rolling resistance occurs when the driving force of the wheels is low[4] and can take a value as low as 75–85% of that occurring in free rolling conditions. The fact that the rolling resistance decreases initially with the application of a driving force, increasing sharply when the force increases, would favor four wheel drive layouts in which all wheels work under moderate driving forces instead of those in which some wheels are idle while others work with higher driving loads.

2.5 STATIC FORCES

Consider a small portion of the tire-road contact area. The force per unit area exerted by the tire on the road can be broken down into a component perpendicular to the road and a tangential component. The first is the contact pressure

[4]D.J. Schuring, *Energy Loss of Pneumatic Tires Under Freely Rolling, Braking and Driving Conditions*, Tire Science and Technology, TSTCA, Vol. 4, No. 1, Feb. 1976, pp. 3–15.

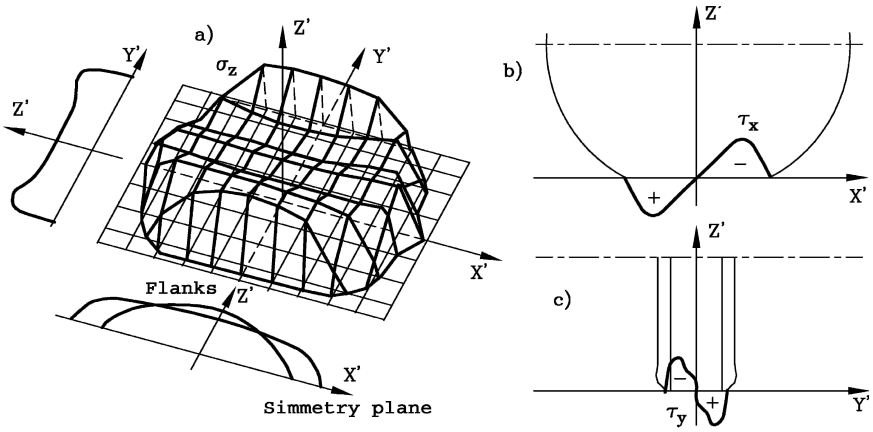


FIGURE 2.27. Force distribution in contact patch. (a) Normal pressure σ_z. (b) Tangential pressure τ_x in the symmetry plane. (c) τ_y and in the $Y'Z'$ plane.

σ_z while the other can be further broken down in the direction of the X' and Y' axes, resulting in the components τ_x and τ_y. The resultant of the distributions of σ_z, τ_x and τ_y are the already defined normal, longitudinal and lateral forces F_z, F_x and F_y respectively.

These distributions are not constant and are strongly influenced by factors such as tire structure, load, inflation pressure, etc. Some typical results obtained on a stationary wheel which exerts no force in the $X'Y'$ plane are reported in Fig. 2.27. At the centre of the contact, the contact pressure σ_z is close to that of the air in the tire at the centre of the contact area, while at the sides it is higher. If the wheel is not rolling the distribution is symmetrical with respect to the $Y'Z'$ plane and the resultant passes through the centre of the contact. Tangential forces do not vanish locally even when no force is exerted in the $X'Y'$ plane, i.e. when their resultant is zero. In such a case components τ_x are directed towards the centre of the contact area and the tire acts to *compact* the ground towards the centre of the contact. Component τ_y has the effect of *stretching* the ground outward.

The force-deflection characteristics of tires depend on many factors, such as travelling speed, pressure and wear. A strong difference exists on this point between bias-ply and radial tires, the latter being less stiff at standstill in every direction.

The characteristics in direction perpendicular to the ground (force F_z versus deflection Z') for some tires are shown in Fig. 2.28. In Fig. 2.28a the curve obtained when removing the load is shown along with that related to the application of the load. A hysteresis cycle can be clearly identified, denoting the presence of damping in the motion along the Z'-axis. This damping is usually at its maximum at standstill and decreases with rolling speed: The practice of

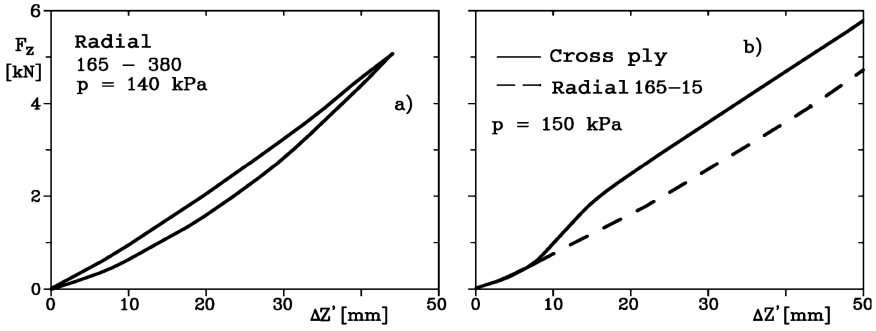


FIGURE 2.28. Contact force F_z as a function of deflection $\Delta Z'$ for some tires in static conditions (load applied and relieved slowly).

neglecting it in simulations of the motion of the vehicle is justified by this observation. In Fig. 2.28b the curves obtained for a radial and a bias-ply tire are compared.

A static tire rate can be defined as the tangent stiffness in any given equilibrium condition, i.e. at any given value of the load, inflation, pressure etc.

Similar plots can be obtained for forces in the X' and Y' directions and moments about the Z' axis versus the corresponding displacements of the centre of the wheel (Fig. 2.29). In all cases the plots show a nonlinear behavior and a hysteresis cycle; radial tires are generally less stiff than bias-ply tires of similar size. All characteristics are influenced by both the speed of rolling and the frequency of application of the force.

2.6 LONGITUDINAL FORCE

Consider a pneumatic wheel rolling on level road. If a braking moment M_b is applied to it, the distributions of normal pressure and longitudinal forces which result from that application are those qualitatively sketched in Fig. 2.30a. The tread band is circumferentially stretched in the zone that precedes the contact with the ground, while in free rolling the same part of the tire is compressed.

The peripheral velocity of the tread band in the leading zone of the contact $\Omega R'_e$ is consequently higher than that (ΩR) of the undeformed wheel. The effective rolling radius R'_e, whose value R_e in free rolling was between R_l and R, grows towards R and, if M_b is large enough, becomes greater than R.

The instantaneous centre of rotation is consequently located below the road surface (Fig. 2.31). The angular velocity Ω of the wheel is lower than that characterizing free rolling in the same conditions ($\Omega_0 = V/R_e$). In such conditions it is possible to define a *longitudinal slip* as:

$$\sigma = \frac{\Omega}{\Omega_0} - 1 = \frac{v}{V} \,, \tag{2.19}$$

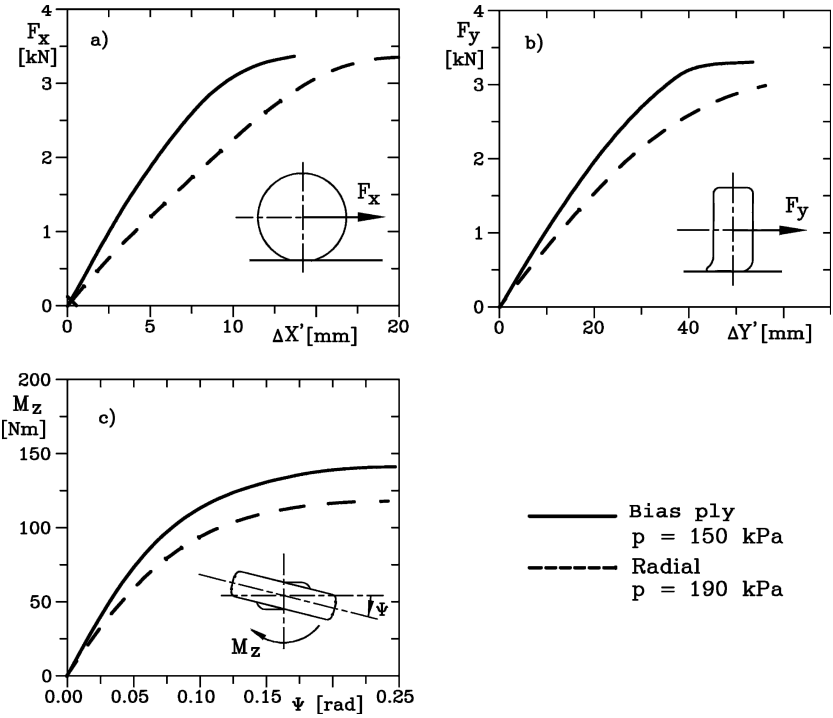


FIGURE 2.29. Forces in the X' and Y' directions and moments about the Z' axis versus deflections $\Delta X'$ and $\Delta Y'$ and rotation about the Z' axis for a radial and a bias-ply tire at standstill (wheel non-rolling) and in static conditions (load applied and removed slowly).

where v is the linear speed at which the contact zone moves on the ground. The longitudinal slip is often expressed as a percentage; in the present book, however, the definition of Eq. (2.19) will be strictly adhered to.

If instead of braking, the wheel is driving, the leading part of the contact zone is compressed instead of being stretched (Fig. 2.30b). The value of the effective rolling radius R'_e is smaller than that characterizing free rolling and is usually smaller than R_l; the angular velocity of the wheel is greater than Ω_0.

The slip defined by Eq. (2.19) is positive for driving conditions and negative for braking.

The presence of the slip velocity[5] v does not mean, however, that there is an actual sliding of the contact zone as a whole. The peripheral velocity of the leading part of that zone is actually $V = \Omega R'_e$. Consequently in that zone no sliding can occur. The speed of the tread band begins to decrease (in braking,

[5]The slip velocity is defined by SAE Document J670 as $\Omega - \Omega_0$, i.e. the difference between the actual angular velocity and the angular velocity of a free rolling tire. Here a definition based on a linear velocity rather than an angular velocity is preferred: $v = R_e(\Omega - \Omega_0)$.

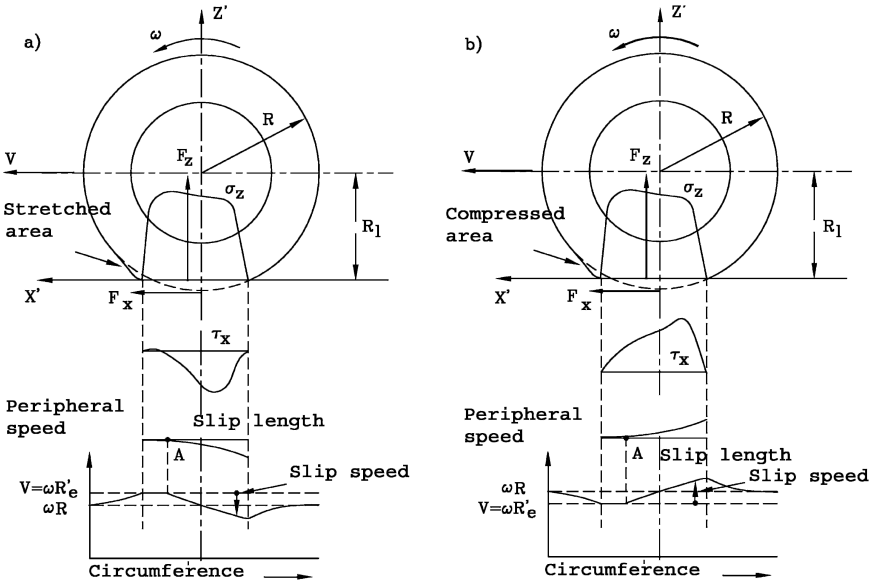


FIGURE 2.30. Contact pressure distribution for a braking (a) and driving (b) wheel. The rolling radius R'_e is different from that of pure rolling R_e (force F_x lies on ground).

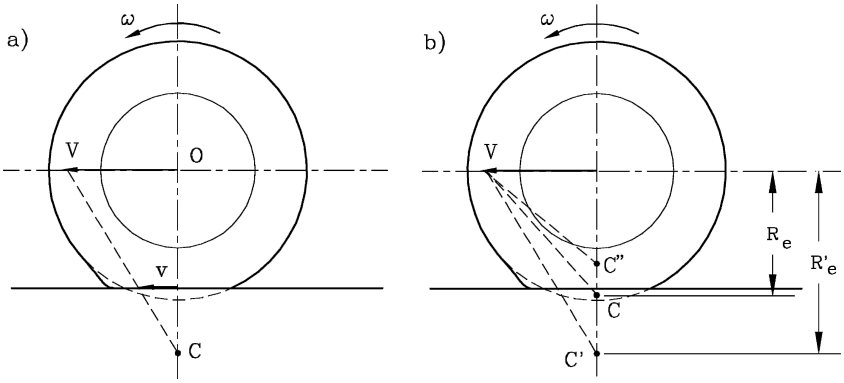


FIGURE 2.31. (a) Braking wheel, center of instantaneous rotation and slip speed. (b) Position of the instantaneous rotation center by pure rotation C, by braking C′ and by traction C″.

increasing in driving) and sliding begins only at the point indicated in Fig. 2.30 as point A. The slip zone, which interests only a limited part of the contact zone for small values of σ, becomes larger with increasing slip and, at a certain value of that parameter, reaches the leading part of the contact zone. Global sliding of the tire then occurs (Fig. 2.32a).

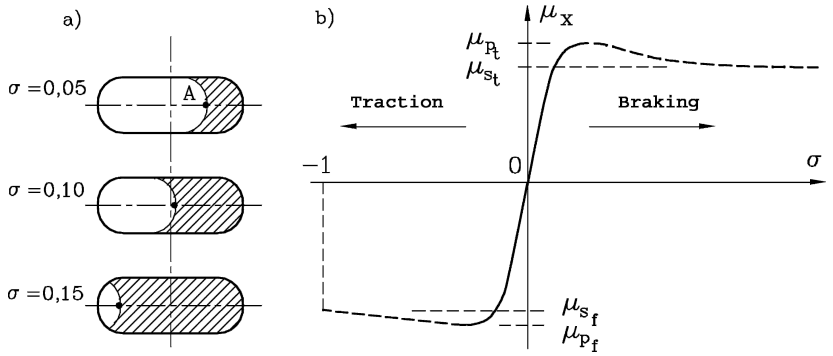


FIGURE 2.32. (a) Slipping area at different values of slip σ. (b) Qualitative diagram of μ_x as function of longitudinal slip σ.

The longitudinal force F_x the wheel exchanges with the road is a function of σ. It vanishes when $\sigma = 0$ (free rolling conditions)[6] to increase almost linearly for values of σ from about -0.25 to about 0.25:

$$F_x = C_\sigma \sigma \ , \tag{2.20}$$

where constant

$$C_\sigma = \left(\frac{\partial F_x}{\partial \sigma} \right)_{\sigma=0}$$

can be defined as *slip stiffness* or *longitudinal stiffness* of the tire.

Outside this range, which depends on many factors, its absolute value decreases in braking up to the value $\sigma = -1$, which characterizes free sliding (locking of the wheel). In driving the force decreases above the stated range, but σ can have any positive value, up to infinity when the wheel spins while the vehicle is not moving.

As a first approximation, force F_x can be considered as roughly proportional to the load F_z, at an equal value of σ. It is consequently useful to define a *longitudinal force coefficient*:

$$\mu_x = \frac{F_x}{F_z} \ . \tag{2.21}$$

The qualitative trend of such a coefficient is reported against σ in Fig. 2.32b.

In the first part of the curve $F_x(\sigma)$, the slip stiffness is roughly proportional to the load and it is possible to define a slip stiffness coefficient $\partial C_\sigma / \partial F_z$, at least for limited variation of the load about a given value F_{z0}:

$$C_\sigma = C_{\sigma 0} + \frac{\partial C_\sigma}{\partial F_z} (F_z - F_{z0}) \, . \tag{2.22}$$

[6]Actually free rolling is characterized by a small negative slip, corresponding to the rolling resistance. This is however usually neglected when plotting curves $F_x(\sigma)$.

Two important values of μ can be identified on the curve both in braking and in driving: The peak value μ_p and the value μ_s which characterizes pure sliding. The first is referred to as the *driving traction coefficient* when the wheel is exerting a positive longitudinal force and as the *braking traction coefficient*, which is usually reported in absolute value, in the opposite case. The second is referred to as the sliding driving traction coefficient and the sliding braking traction coefficient respectively.

The part of the curve $\mu(\sigma)$ which lies beyond the range included by the two peak values, represented by a dashed line in Fig. 2.32b , is a zone of instability in the practical use of the vehicle. The equation of motion of a braking wheel is:

$$J\frac{d\Omega}{dt} = |F_x|R_l - |M_f| \; . \tag{2.23}$$

If a decrease of σ at constant speed V leads to a decrease of the absolute value of μ_x, it causes a decrease of the absolute value of F_x, i.e. of the force which maintains the rotation of the wheel. If this decrease of $|F_x|$ is not quickly followed by a decrease of the braking moment $|M_f|$ (and it is unrealistic to assume that the driver can react fast enough to release the brakes) a further slowing down of the wheel takes place, which in turn causes a further reduction of $|F_x|$.

Actually, when the optimum value of σ, which is characterized by μ_{p_b}, is exceeded, the wheel locks in a very short time. In order to prevent the locking of wheels, devices generally defined as anti-lock or anti-skid systems are now widely used.

Such devices detect the deceleration of the wheel and, when it reaches a predetermined value, react by decreasing the braking moment and avoiding the locking of the wheel. *Anti-lock* devices can operate on each wheel separately or, more often, on both wheels of an axle. Similarly, to prevent a wheel from slipping under the effect of an applied driving torque, *anti-spin* devices limit the engine moment when the acceleration of the driving wheels exceeds a stated value.

The curves usually show a certain symmetry between braking and driving conditions and the maximum braking and driving forces are often assumed to be equal. The values of function $\mu_x(\sigma)$ depend on a number of parameters, such as the type of tire, road conditions, speed, magnitude of the side force F_y exerted by the tire and many others. Moreover, there is a significant difference between the curves obtained by different experimenters in conditions not exactly comparable. Some curves $\mu_x(\sigma)$ obtained in different conditions are shown in Fig. 2.33.

The maximum value of the longitudinal force decreases with increasing speed but this reduction is greatly influenced by operating conditions. Generally speaking, it is not marked on dry road, while it is greater on wet surfaces. Moreover, the difference between the maximum value (see Table 2.5) and the value related to sliding (Table 2.6) is more notable on wet roads (Fig. 2.34). Particularly dangerous conditions are encountered when the road is only partially wet and dirty: The behavior can vary from spot to spot and the value at slip can be very much different from the maximum.

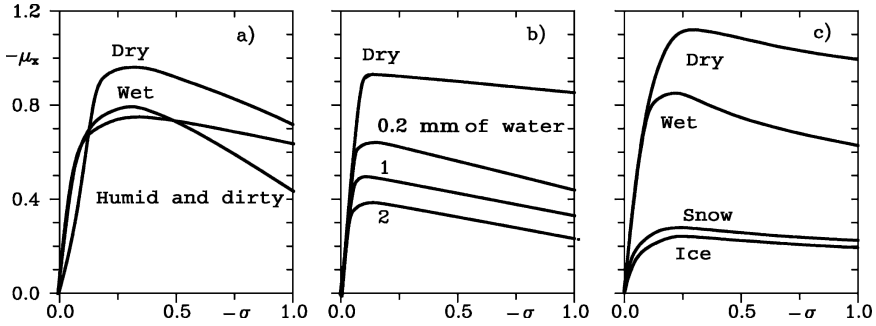


FIGURE 2.33. Curves of $\mu_x(\sigma)$ obtained in different conditions.

TABLE 2.5. Values of μ_p for different tires at 30 km/h. Tires 145 15; 150 15 e 165 R 15; $F_z = 3$kN, $p = 160$ kPa for cross ply tires (conv.) and 220 kPa for radial tires.

Tire	Road					
	Concrete		Tarmac			
	Dry	Wet	Dry	Wet	Snow	Ice
Rad.	1.19	0.99	1.22	1.10	0.45	0.25
Conv.	1.13	0.84	1.02	1.07	0.27	0.24
Rad. (snow)	1.04–1.12	0.62–0.83	1.00–1.09	1.00–1.10	0.36–0.47	0.24–0.44
Conv. (snow)	0.86–1.02	0.59–0.70	0.81–0.89	0.78–1.02	0.41–0.48	0.29–0.37
Chains					0.60	0.40

TABLE 2.6. Values of μ_s for the tires of the previous table.

Tire	Road					
	Concrete		Tarmac			
	Dry	Wet	Dry	Wet	Snow	Ice
Rad.	0.95	0.73	1.03	0.90	0.43	0.16
Conv.	0.99	0.62	0.88	0.80	0.22	0.18
Rad. (snow)	0.88–1.00	0.50–0.61	0.87–0.99	0.77–0.93	0.35–0.45	0.22–0.41
Conv. (snow)	0.72–0.90	0.47–0.57	0.70–0.78	0.67–0.84	0.39–0.47	0.29–0.36

The values shown in the tables must be regarded only as average indications as they are quite sensitive. Note that in good conditions the longitudinal force can be almost equal to the load acting on the tire or even greater; the values reported, however, refer to standard tires used on passenger vehicles. High performance tires, particularly those used on racing cars, show peak values of μ_x which can be as high as $1.5 \div 1.8$, but even these tires do not reach very high values of longitudinal force coefficient in sliding conditions; the difference between μ_p and μ_s is even larger. To reach the highest level of performance, particular formulations of the rubber must be used. These are characterized by strong wear and consequently are restricted to competition tires. Note that radial tires almost always show better performances than bias-ply types.

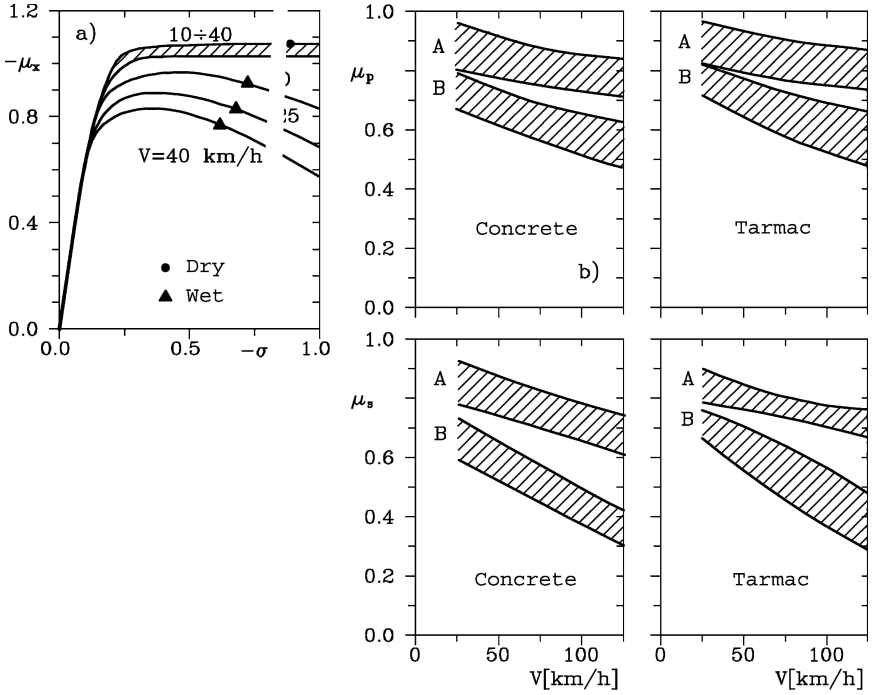


FIGURE 2.34. Influence of car speed on μ_p and μ_s on wet and dry road. Cross and bias ply tires 165 l5 e 150 R l5 with $F_z = 3{,}75$ kN.

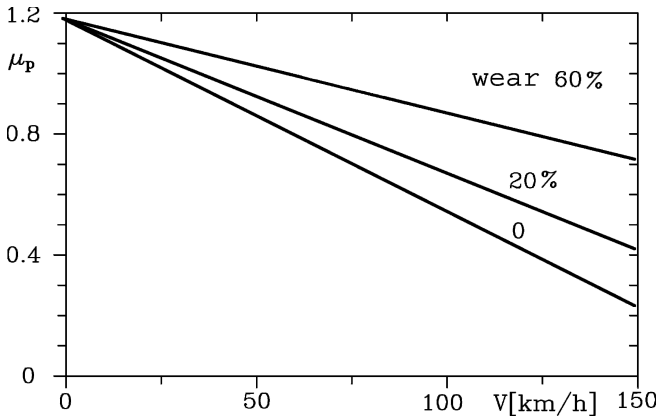


FIGURE 2.35. Influence of tread wear on the peak value of longitudinal friction as a function of speed.

Tread wear has a great influence on longitudinal forces, particularly at high speed. From Fig. 2.35 it can be assumed that the increase of μ_{p_b} due to wear can be, particularly at high speed, quite noticeable. It must, however, be noted that

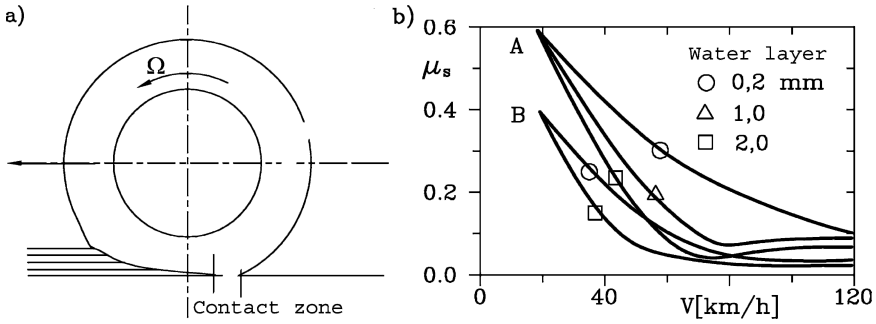


FIGURE 2.36. Hydrodynamic lift of a wheel (*aquaplaning*). (a) Scheme; (b) diagram of μ_s as a function of speed on a wet road. Tire 145 - l5 with tread (curve A) and without (curve B); $F_z = 3$ kN; $p = 150$ kPa.

the figure refers to dry roads, as the presence of even a thin water layer on the road can drastically change these results.

When the road is wet, particularly if the water layer is thick, the tire can lift from the road surface as a result of hydrodynamic lift (*aquaplaning*). A thin layer of water can slip between the tire and the road, thus reducing the contact area (Fig. 2.36). With increasing speed the area of the contact zone further reduces, until a complete lifting of the tire takes place. True hydrodynamic lubrication conditions can be said to exist in this case and consequently the force coefficient (or, better, the friction coefficient as in this condition sliding usually occurs) reduces to low values, on the order of 0.05.

In order to avoid hydroplaning, or at least to postpone its occurrence, it is mandatory to displace water from the contact zone as quickly and effectively as possible. This can be done in two distinct ways: By making the road surface permeable or by using deep grooves in the tread in both circumferential and transversal direction in order to permit a high flow rate.

The assumption, in a way implicit in the definition of the longitudinal force coefficient, that longitudinal forces are proportional to the normal force acting on the wheel is only a crude approximation. Actually the longitudinal force coefficient decreases with increasing load, as shown in Fig. 2.37.

The curves $\mu_x(\sigma)$ can be approximated by analytical expressions. One of the formulas that can be used in the range $-1 < \sigma < 1$ is:

$$\mu_x = A\left(1 - e^{-B\sigma}\right) + C\sigma^2 - D\sigma \,, \tag{2.24}$$

where:

$$B = \left(\frac{K}{\alpha + d}\right)^{1/n}$$

is a factor which takes into account the interaction between the longitudinal slip σ and the sideslip α (see Sect. 2.8). The derivative in the origin $(\partial\mu_x/\partial\sigma)_{\sigma=0}$ is simply $AB - D$. Coefficients A, C, D, K, d and n must be obtained from the experimental curves and have no physical meaning. They depend not only on

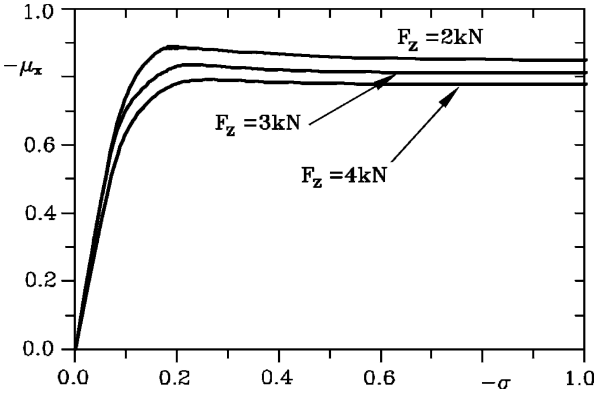


FIGURE 2.37. Influence of the vertical load F_z on function $\mu_x(\sigma)$. Tire 150 15, $p = 170$ kPa, $V = 100$ km/h.

the road conditions but also on load. A curve $\mu_x(\sigma)$ for a radial tire 145/80 R 13 4.5J obtained through Eq. (2.24) is reported in Fig. 2.38a, curve A.

An excellent approximation of longitudinal force F_x as a function of the slip σ can be obtained through the empirical equation introduced by Pacejka[7] and known as the *magic formula*. Such a mathematical expression allows one to express forces F_x and F_y and the aligning torque M_z as functions of the normal force F_z, of σ and of the sideslip (α) and camber (γ) angles.

The equation yielding the longitudinal force F_x, as a function of the slip σ is:

$$F_x = D \sin\left(C \arctan\left\{B(1-E)(\sigma+S_h)+E \arctan\left[B(\sigma+S_h)\right]\right\}\right)+S_v \ , \quad (2.25)$$

where B, C, D, E, S_v and S_h are six coefficients that depend on the load F_z and on angle γ. These must be obtained from experimental testing and have no direct physical meaning. In particular, S_v and S_h have been introduced to allow non-vanishing values of F_x when $\sigma = 0$.

Coefficient D yields directly the maximum value of F_x, apart from the effect of S_v. The product BCD gives the slope of the curve for $\sigma + S_h = 0$. The values of the coefficients are expressed as functions of a number of coefficients b_i, which can be considered as characteristic of any specific tire, but depend also on road conditions and speed:

$$C = b_0 \qquad D = \mu_p F_z \ ,$$

[7]E. Bakker, L. Lidner, H.B. Pacejka, *Tire Modelling for Use in Vehicle Dynamics Studies*, SAE Paper 870421; E. Bakker, H.B. Pacejka, L. Lidner, *A New Tire Model with an Application in Vehicle Dynamics Studies*, SAE Paper 890087.

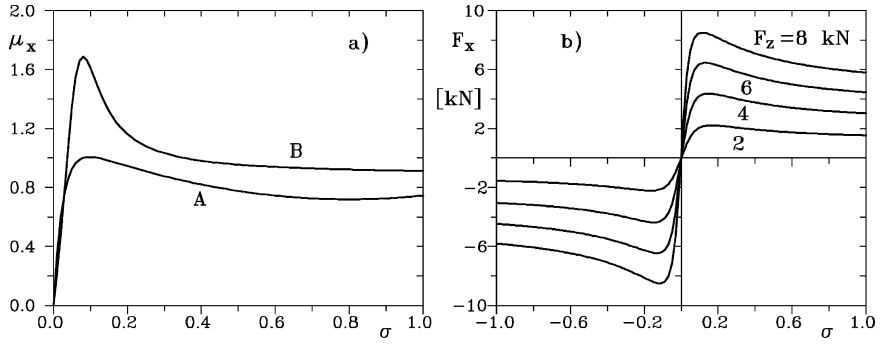


FIGURE 2.38. (a) Curves $\mu_x(\sigma)$ for a 145 R 13 tire obtained with Eq. (2.24) (curve A) and for a 245/65 R 22.5 tire obtained from Fig. (2.25) (curve B). (b) Curves $F_x(\sigma)$ for some values of vertical load Fig. (2.25) for a 205/60 R 15 V tire.

where for b_0 a value of 1.65 is suggested and:

$$\mu_p = b_1 F_z + b_2 \ , \qquad BCD = \left(b_3 F_z^2 + b_4 F_z\right) e^{-b_5 F_z} \ ,$$
$$E = b_6 F_z^2 + b_7 F_z + b_8 \ , \quad S_h = b_9 F_z + b_{10} \ , \qquad S_v = 0 \ .$$

Note that product BCD is nothing other than the slip stiffness of the tire.

If a symmetrical behavior for positive and negative values of force F_x is accepted, this model can be used for both braking and driving. The curve is usually extended to braking beyond the point where $\sigma = -1$, to simulate a wheel rotating in the backward direction while moving forward.

The coefficients introduced in Eq. (2.25) and the results obtained from it are usually expressed in inconsistent units: Force F_z is in kN, longitudinal slip is expressed as a percentage and force F_x is in N.

A set of curves $F_x(\sigma)$ obtained for vertical loads F_z = 2, 4, 6 and 8 kN for a radial tire 205/60 VR 15 6J is shown in Fig. 2.38b.

A curve $\mu_x(\sigma)$ for a high performance radial tire 245/65 R 22.5 obtained from Eq. (2.25), is also shown in Fig. 2.38a, curve B. Note the high peak value of the force coefficient.

Using the same equations, in Fig. 2.39 the curve $F_x(\sigma)$ is shown for a tire 195/65 R 15: Note the flat shape of the curve, after the peak, due to the high performance type of rubber mix.

The importance of the model expressed by Eq. (2.25) is primarily linked to the fact that tire manufacturers increasingly disclose the performance of their tires in terms of magic formulae coefficients. If this trend consolidates, the magic formula will prove to be a simple and accurate model for tire behavior and, even more important, one for which the data will be readily available.

Magic formulae were modified by the Author in 1996 by the introduction of additional parameters; the new formulae include modelling the combined effect of longitudinal slip and sideslip, as we will see in a dedicated paragraph.

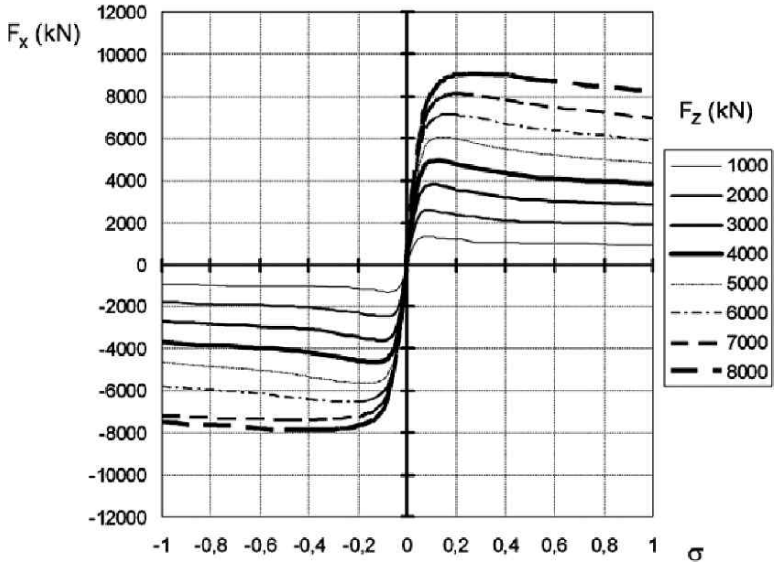


FIGURE 2.39. Curves $F_x(\sigma)$ for a 195/65 R 15 tire obtained for different values of vertical load F_z.

2.7 CORNERING FORCES

In the previous section it became clear that a pneumatic tire can exert longitudinal forces only if deformations are present in the tread band and if the wheel has a non-vanishing longitudinal slip. In the same way the generation of cornering forces cannot be understood unless reference is made to the lateral deformations of the tire and its sideslip angle: The generation of tangential forces in the road-wheel contact is directly linked with the compliance of the tire.

The fact that the wheel has a sideslip angle, i.e. does not engage in pure rolling, does not mean that in the contact zone the tire slips on the road: As seen for longitudinal forces, the compliance of the tire allows the tread to move, relative to the centre of the wheel, with the same velocity as the ground. Some localized sliding between the wheel and the road can be present, however, and, with increasing sideslip angle, becomes more and more important, until the whole wheel is engaged in actual, macroscopic sliding.

If the velocity of the centre of the wheel does not lie in its mean plane, i.e. if the wheel travels with a sideslip angle, the shape of the contact zone is quite distorted (Fig. 2.40). Consider a point belonging to the mean plane on the tread band. Upon approaching the contact zone it tends to move in a direction parallel to the velocity V, relative to the centre of the wheel, and consequently goes out of the mean plane.

After touching the ground at point A, it continues following the direction of the velocity V (for a stationary observer, it remains still) until it reaches

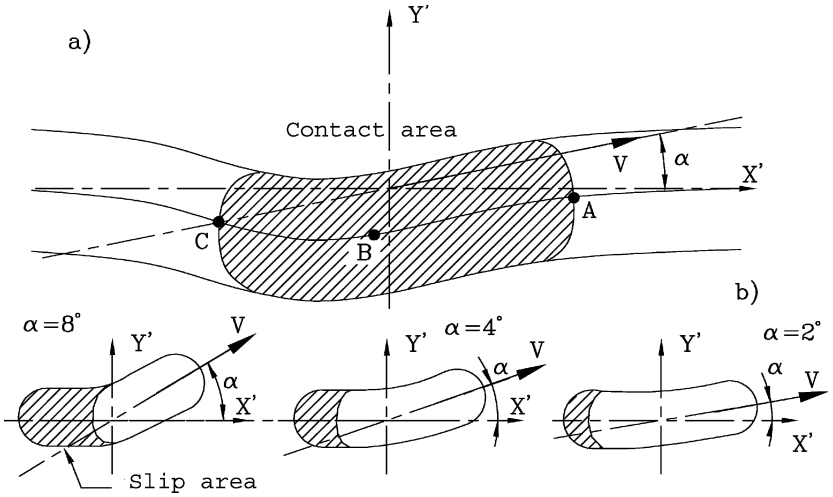


FIGURE 2.40. Wheel-road contact when sideslip angles are present. (a) Contact zone and path of a point of the tread on the equator plane; (b) contact zone and slip zone at different values of α (α is not in scale).

point B. At that point, the elastic forces pulling it towards the mean plane are strong enough to overcome those due to the friction on the road, forcing it to slide on the road and to deviate from its path. This sliding continues for the remaining part of the contact zone until point C is reached. The contact zone can thus be divided into two parts: A leading zone in which no sliding occurs and a trailing zone in which the tread slips towards the mean plane. This second zone grows with the sideslip angle (Fig. 2.40b), until it pervades the entire contact zone and the wheel actually slips on the ground.

The lateral deformations of the tire are plotted in a qualitative way in Fig. 2.41, together with the distribution of σ_z, τ_y, and the lateral velocity. The resultant F_y of the distribution of side forces is not applied at the centre of the contact zone but at a point that is located behind it at a distance t. Such a distance is defined as *pneumatic trail*.

The moment $M_z = F_y t$ is the *aligning moment* as it tends to force the mean plane of the wheel towards the direction of the velocity V. The absolute value of the side force F_y grows almost linearly at first as α increases. Then, when the limit conditions of sliding are approached, it grows more slowly. Eventually it remains constant, or decreases slightly, when sliding conditions are reached.

The side force F_y is plotted as a function of α for the cases of a radial and a bias-ply tire in Fig. 2.42a. Radial tires show a stiffer behavior than bias-ply tires in terms of side forces, as they require smaller sideslip angles to produce the same side force.

With increasing sideslip angle, τ_y is more evenly distributed and the pneumatic trail decreases. The aligning moment is consequently the product of a force

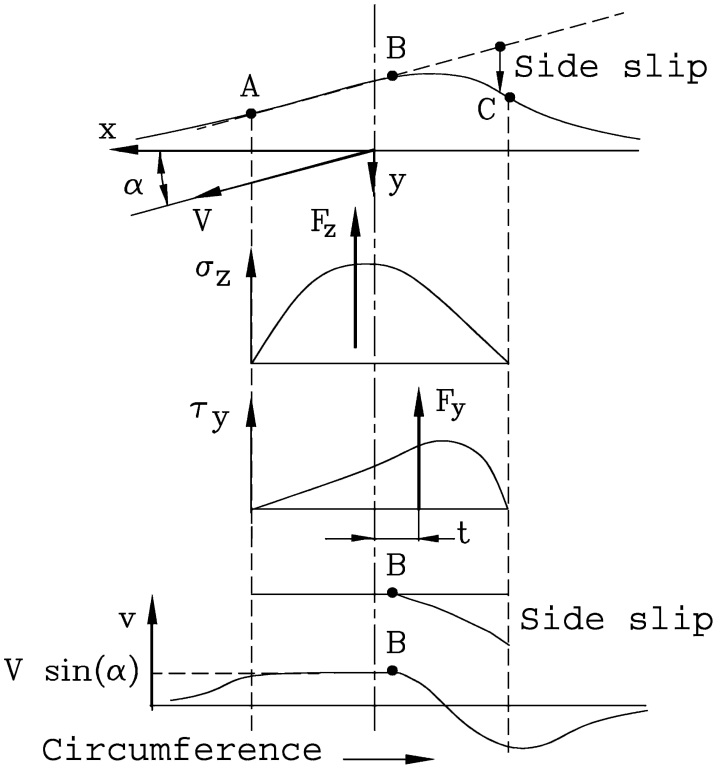


FIGURE 2.41. Lateral deformation, distribution of pressures σ_z e τ_y, slip and lateral speed in a cornering tire.

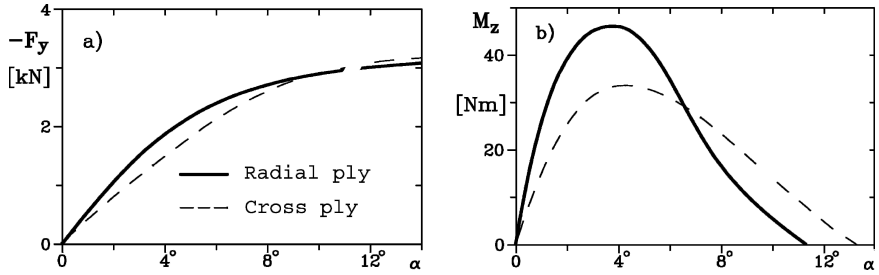


FIGURE 2.42. Lateral force F_y and self aligning moment M_z for different tires of the same size 145 D 13 and 145 R 13; $F_z = 3$ kN, $p = 170$ kPa; $V = 40$ km/h.

that increases with α and a distance that decreases; its trend is consequently of the type shown in Fig. 2.42b. At high values of α, M_z can change direction, as is shown in the figure.

The side force coefficient

$$\mu_y = \frac{F_y}{F_z}$$

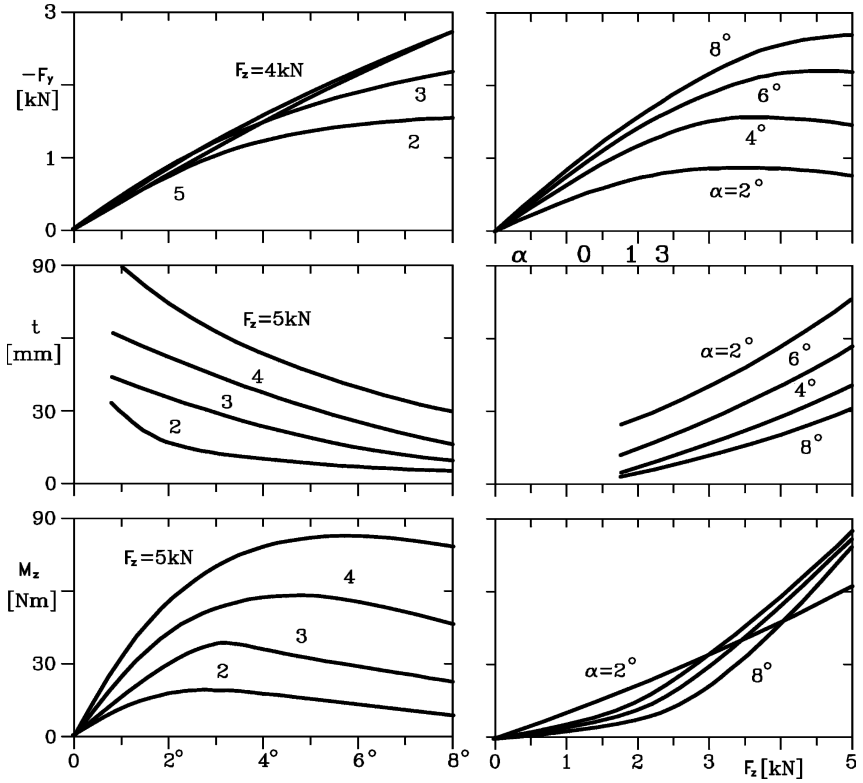


FIGURE 2.43. Diagrams of F_y, M_z and t as function of F_z and α. Tire 145 D 15, $p =$ 180 kPa, $V = 50$ km/h.

is often used. Its maximum value, usually defined as the lateral traction coefficient, is written as μ_{y_p} and the value taken in sliding conditions as μ_{y_s}.

Both force F_y and moment M_z depend on many factors besides the angle α, such as normal force F_z, speed, pressure p, road conditions, etc. Diagrams of cornering forces F_y, of self aligning moment M_z and of trail t are shown in Fig. 2.43 as a function of the sideslip angle α and the vertical load F_z. The figure refers to a 145 D 15 tire, inflated at 180 kPa at a speed of 50 km/h.

It must be observed that all curves are to be taken as correct only insofar as their shape is concerned. They are less precise in terms of numeric values; they depend, in fact, on the kind of test machine, used for measurement, that cannot simulate on-road behavior correctly. See the following paragraph dedicated to test machines.

The diagrams we are going to show cannot substitute for experimental data gathered on real cars, but they offer a trend indicator for the interpretation of experimental results.

The lateral behavior of the tire can be summarized in a single diagram, the so-called *Gough diagram*, in which the side force F_y is plotted against the self

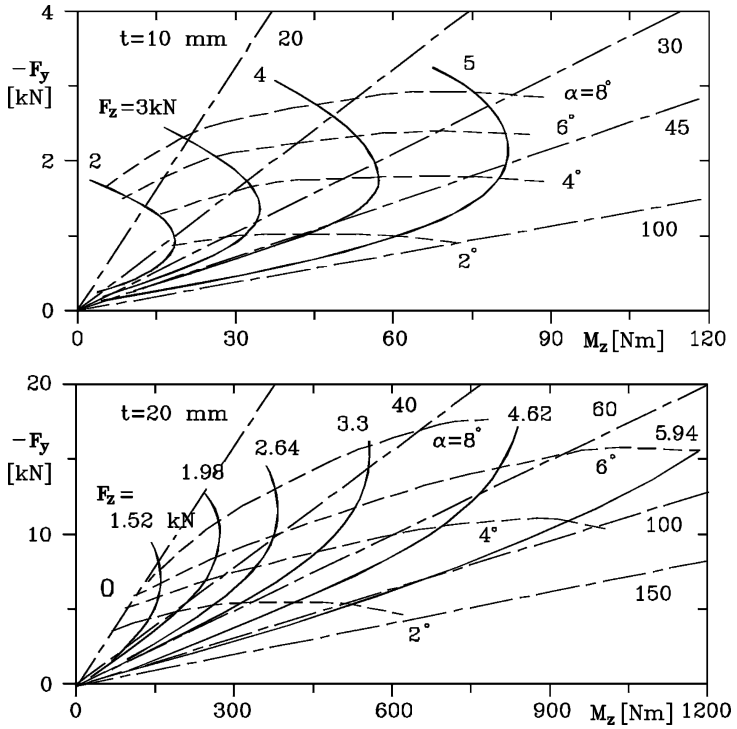


FIGURE 2.44. Gough diagrams. (a) Radial ply tire for a car 145 R 16; $p = 180$ kPa, $V = 50$ km/h; (b) cross ply tire for an industrial vehicle 300 20, $p = 675$ kPa.

aligning torque M_z with F_z, α and t as parameters. The Gough diagrams for two different tires are shown in Fig. 2.44.

Alternatively, the *carpet plot* can be drawn; here the force is plotted against the slip angle and the various curves at different F_z are simply superimposed, translated by a quantity proportional to the normal force with respect to each other. The aligning torque can also be plotted in this way. Carpet plots are shown in Fig. 2.59.

At increasing speed, the curve $F_y(\alpha)$ lowers, primarily in the area corresponding to higher values of the sideslip angle. The linear part remains almost unchanged (Fig. 2.45). Likewise the pneumatic trail t decreases with increasing speed and consequently the aligning torque shows a marked decrease compared to that of the side force.

The decrease of F_y and of M_z as speed increases is more marked on slippery roads, as we can see in Fig. 2.46, where the curves seen in Fig. 2.45 are shown at two different speeds on dry and wet road.

In terms of hydrodynamic lifting (aquaplaning), the same considerations seen for the longitudinal force F_x apply to the side force F_y. The decrease of aligning torque which goes together with the decrease of the side force should

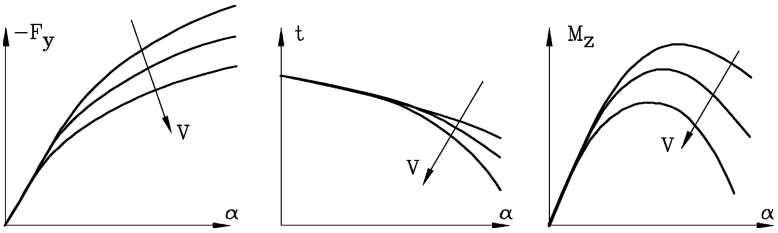


FIGURE 2.45. Qualitative curves of $F_y(\alpha)$, $M_z(\alpha)$ and $t(\alpha)$ at different speeds.

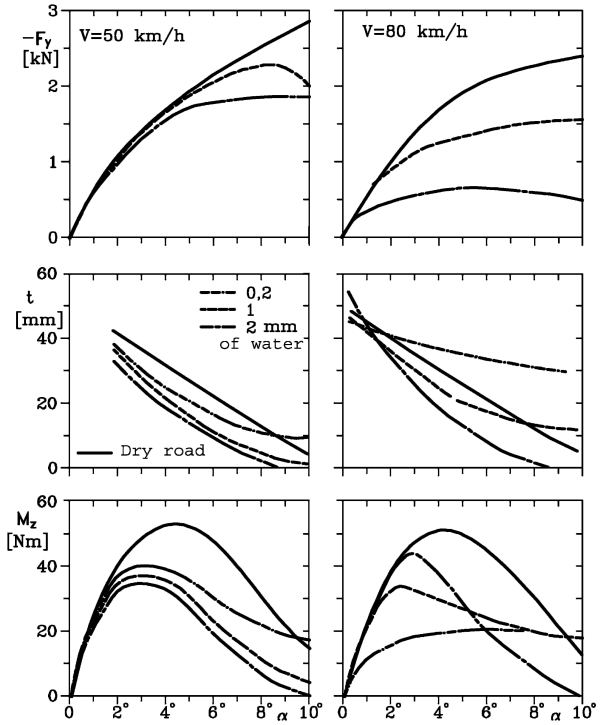


FIGURE 2.46. Curves of $F_y(\alpha)$, $M_z(\alpha)$ and $t(\alpha)$ at two different speeds (50 and 80 km/h) on dry and wet road. Tire 155 R 15, $F_z = 3\,\text{kN}$, $p = 150$ kPa.

warn the driver of the approaching of loss of traction. A poorly designed tire, however, may maintain strong aligning moments even in conditions of reduced transversal traction, leading the driver into error.

Finally, in Fig. 2.47 we show a carpet plot for cornering forces and in Fig. 2.48 a plot for self aligning torque, for two tires 195/65 R 15 and 225/45 R 17 proposed for two versions of the same car.

It should be remembered that even if tire widths are different the free rolling radius of both tires is the same and equal to 317 mm, both tires using the same wheel house.

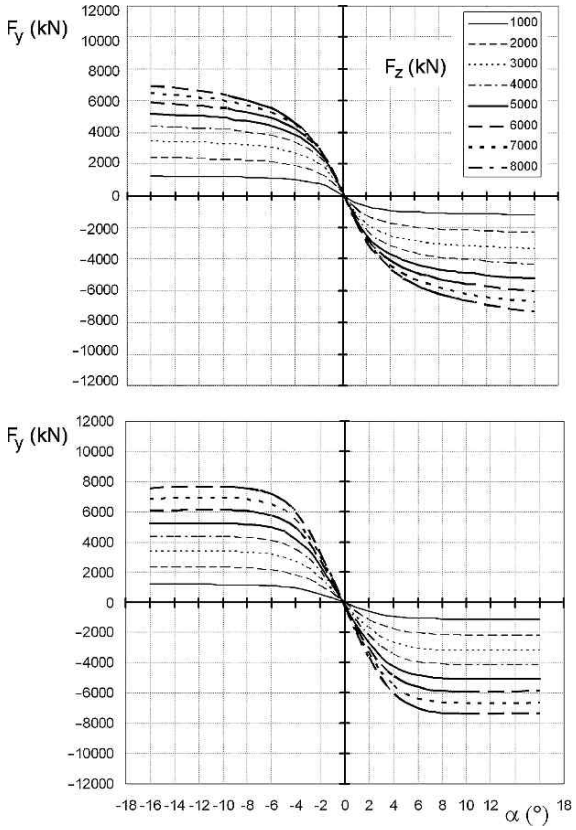


FIGURE 2.47. Carpet diagrams of cornering force for two tires of same diameter but different aspect ratio: 195/65 R 15 and 225/45 R 17. The parameter used is vertical load.

The effect of increased width combined with the different aspect ratio can be seen in cornering stiffness; stiffer tires can be valuable for higher cornering force at the same sideslip angle, as in sport versions. Note that the effect on the self aligning moment is not positive; this result can be corrected with a different king-pin geometry.

The presence of a camber angle, even when no slip angle is present, produces a lateral force. It is usually called *camber thrust* or *camber force*, as distinct from cornering force, which is due to sideslip angle alone. The camber force added to the cornering force gives the total side or lateral force. The camber force is usually far smaller than the cornering force, at least at equal values of angles α and γ. It depends on the load F_z, being practically linear with it (Fig. 2.49), and is strongly dependent on the type of tire under consideration.

The camber thrust is usually applied at a point ahead of the centre of the contact zone, producing a small moment M_{z_γ}. It is usually neglected because of its very small value. Bias-ply tires usually produce greater camber thrust and moment than radials.

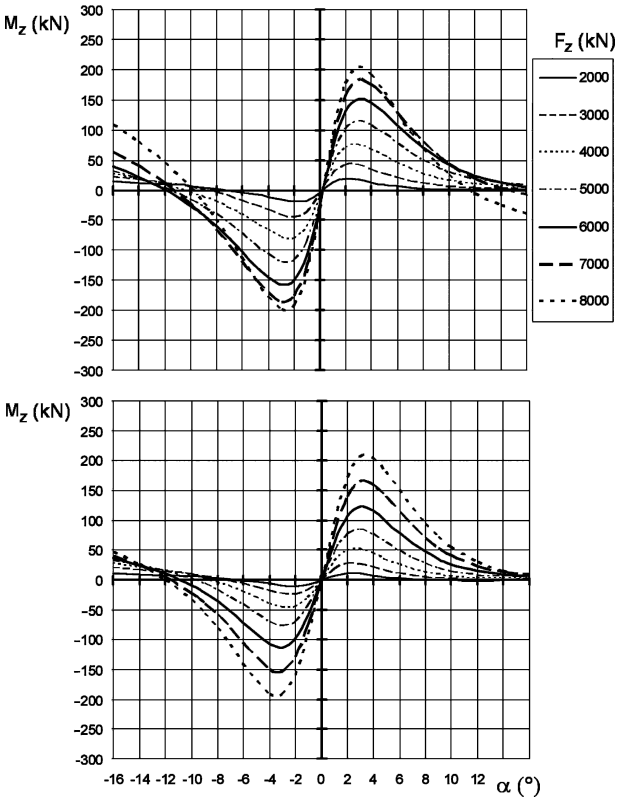


FIGURE 2.48. Carpet diagrams of self aligning torque for two tires of same diameter but different aspect ratio: 195/65 R 15 and 225/45 R 17. The parameter used is vertical load.

Both sideslip and camber are simultaneously present in most cases. As an example, the effect of camber on the force due to sideslip for two different values of the load is shown in Fig. 2.50. The camber thrust is more noticeable at low sideslip than when the sideslip is large, particularly when the wheel is less loaded. The aligning torque under the combined effect of sideslip and camber for a radial tire is shown in Fig. 2.51.

In theory both sideslip angle and camber angle should be zero when cornering force and self aligning torque are zero. This is not always true in practice, as is shown in carpet diagrams in Figs. 2.52 and 2.53, respectively, for the cornering force and the self aligning torque.

Moreover, the centre of the hysteresis cycle is not at the point at which both angle and force are equal to zero: Because of the lack of geometrical symmetry, a tire working in symmetrical conditions produces a side force. A first effect is caused by a possible *conicity* of the outer surface of the tire: A conical drum would roll on a circular path whose centre coincides with the apex of the cone.

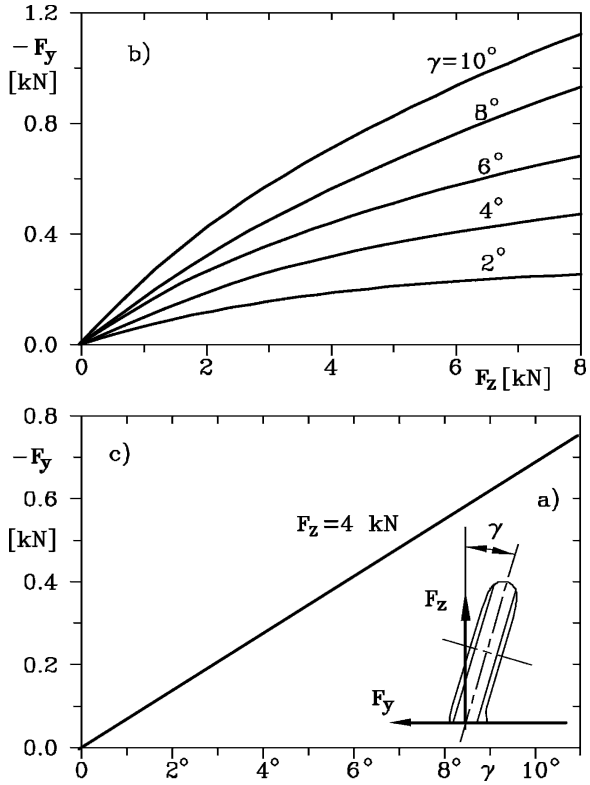


FIGURE 2.49. Camber force. (a) Scheme; note that F_y is negative and directed to the opposite as shown. Camber force as function of vertical load (b) and camber angle (c). Tire 165 13, $p = 200$ kPa.

Conicity is caused by lack of precision during the manufacturing process and hence is linked with manufacturing quality control; its direction is random and its amount changes from tire to tire of the same model. If a tire is turned on its rim, the direction of conicity is reversed, as is the force it creates when the tire is rolled along a straight path.

Another unavoidable lack of symmetry is linked with the angles of the various plies and their stacking order; the effect it causes is called *ply steer*. If the wheel is rolled free, ply steer causes it to roll along a straight line angled with respect to the plane of symmetry; if the wheel rolls with no sideslip angle the generation of a side force results. If a tire is turned on its rim the direction of the force due to ply steer is not reversed. As it is caused by tire design factors, the effect of ply steer, unlike that of conicity, is very consistent between tires of the same model.

While conicity can be included into tire models only in a statistical way, ply steer is one of the peculiarities of each tire and can be accounted for with precision. Note that while these effects are usually considered a nuisance, opposite

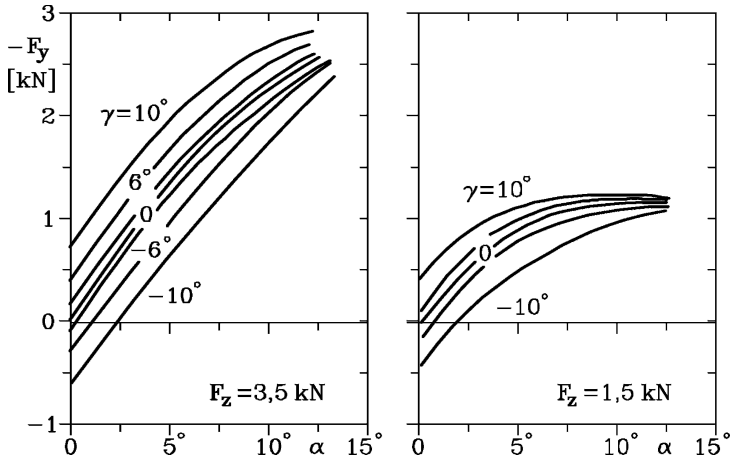


FIGURE 2.50. Cornering force as a function of the sideslip angle for different values of camber angle and for two values of the vertical force. Tire 135 13, $p = 160$ kPa, $V = 40$ km/h.

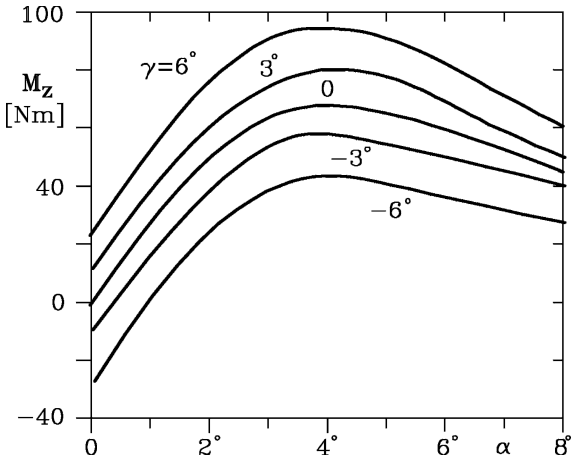


FIGURE 2.51. Self aligning torque as a function of sideslip angle, for different camber angles. Radial tire 185 R 14, $p = 230$ kPa, $F_z = 3.5$ kN.

ply steer of the wheels of a given axle can even be used as a substitute for toe-in; while the latter increases, the rolling resistance of the first has no effect on it.

Generally speaking, the lateral force offset is subdivided into two parts: The part that does not change sign when the direction of rotation is reversed is said to be ply-steer force, while the part that changes sign is said to be conicity force.

As already stated, for low values of the sideslip angle the cornering force increases linearly with α

$$F_y = C_\alpha \alpha \, . \tag{2.26}$$

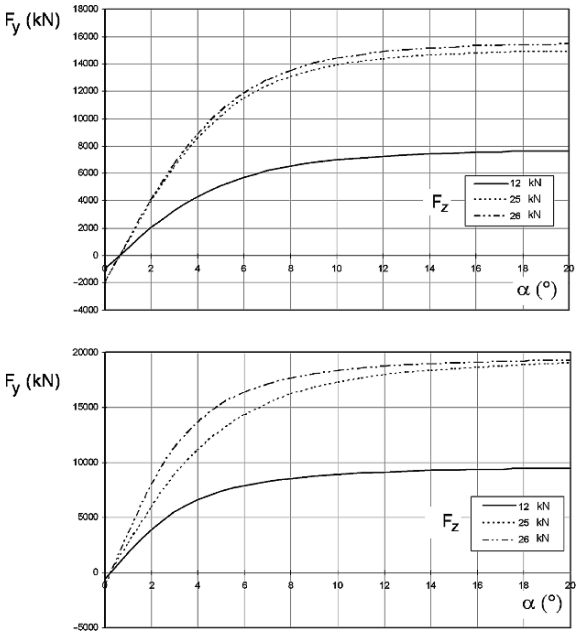


FIGURE 2.52. Carpet diagrams for cornering force as a function of sideslip angle, for two industrial vehicle tires 315/80 R 22,5; the stiffer has a reinforced belt for steering wheels (single wheel), while the other is for driving wheels (twin wheels). The parameter is the vertical force.

Coefficient

$$C_\alpha = \left(\frac{\partial F_x}{\partial \alpha} \right)_{\alpha=0}$$

is always negative, because a positive sideslip angle causes a negative side force. The *cornering stiffness* or *cornering power* C is usually expressed as a positive number, so it is arbitrarily defined and written as the derivative $\partial F_y/\partial\alpha$ changed in sign, and

$$F_y = -C\alpha \; . \tag{2.27}$$

This relationship obviously holds only for small enough values of α.

Expression (2.27) is quite useful for studying the dynamic behavior of vehicles under the assumption of small sideslip angles, as it actually occurs in normal driving conditions. In particular, it is essential in the study of the stability of linearized models.

Figure 2.54 shows a diagram of this parameter, as a function of vertical load and sideslip angle for the 315/80 R 22,5 tires previously described in carpet plots.

The ratio between the cornering stiffness and the normal force is usually referred to as the *cornering stiffness coefficient* (the term cornering coefficient is also used but SAE recommendation J670 suggests avoiding it for clarity). For

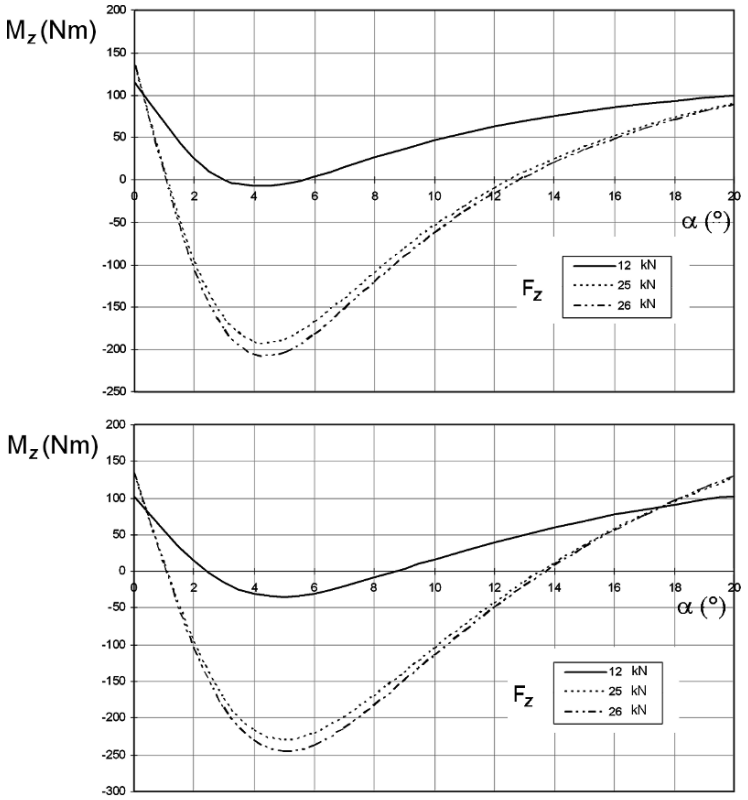


FIGURE 2.53. Carpet diagram of self aligning torque as function of sideslip angle for the same industrial vehicle tires 315/80 R 22,5 of previous figure.

bias-ply tires it is of the order of 0.12 $\text{deg}^{-1} = 6.9\ \text{rad}^{-1}$ and for radial tires of the order of 0.15 $\text{deg}^{-1} = 8.6\ \text{rad}^{-1}$. As the cornering coefficient changes greatly from tire to tire, it is interesting to note the distribution of its value for tires of different types (Fig. 2.55a).

In the same way the camber stiffness can be defined as the slope of the curve $F_y(\gamma)$ for $\gamma = 0$:

$$C_\gamma = \left(\frac{\partial F_y}{\partial \gamma} \right)_{\gamma=0} .$$

Note that the camber thrust produced by a positive camber angle is negative and hence the camber stiffness is negative. The ratio between the camber stiffness and the normal force is usually referred to as the *camber stiffness coefficient*. This coefficient is higher for bias-ply tires than for radial tires: In the first case an average value is on the order of 0.021 $\text{deg}^{-1} = 1.2\ \text{rad}^{-1}$ and in the second is of the order of 0.01 $\text{deg}^{-1} = 0.6\ \text{rad}^{-1}$. The distribution of the camber coefficient for tires of different types is also of interest (Fig. 2.55b).

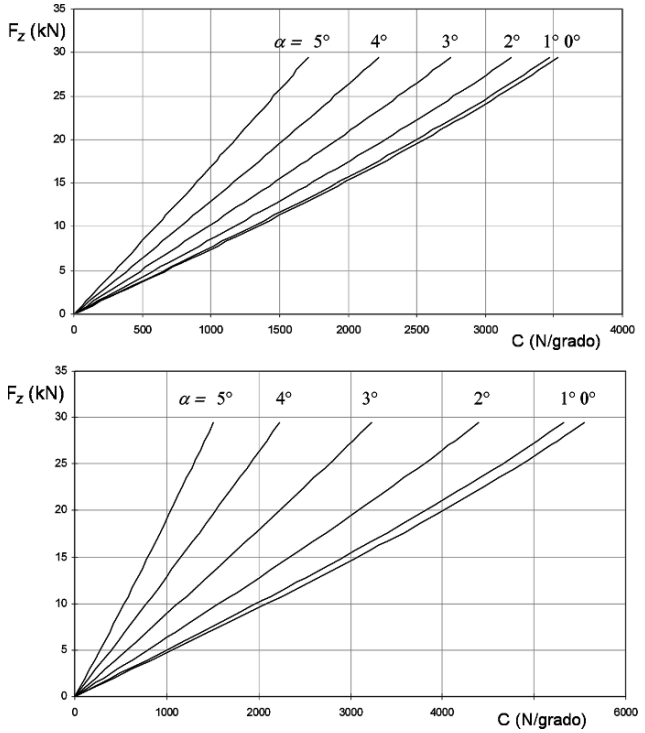


FIGURE 2.54. Carpet plot of vertical force as a function of the cornering stiffness for the two tires of the previous figures. The parameter is the sideslip angle.

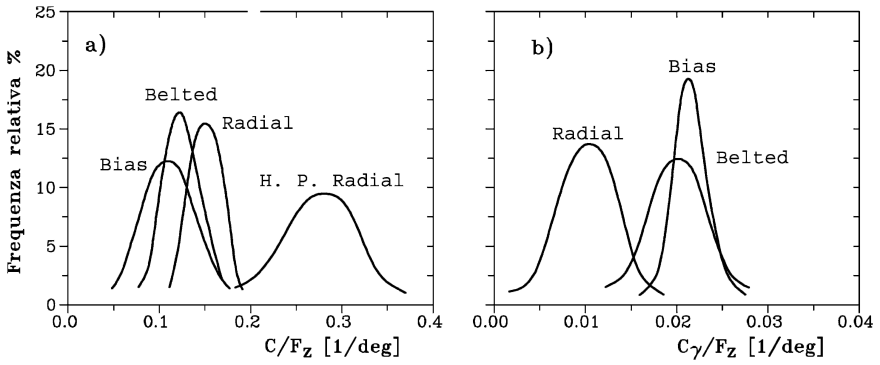


FIGURE 2.55. Frequency distribution of the cornering stiffness (a) and camber stiffness (b) for different car tire types.

The value of the camber stiffness is important in cases a where wheel rolls on a road with a transversal slope with its mid plane remaining vertical: In this case there is a component of the weight which is directed downhill, while the camber

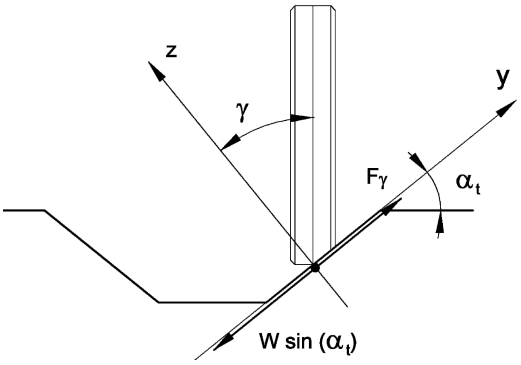


FIGURE 2.56. Wheel on a road with lateral slope α_t: camber force and weight component along the slope.

thrust is directed uphill. The net effect can appear in one direction or the other depending on the magnitude of the camber stiffness coefficient (Fig. 2.56).

The downhill component of the weight is $W \sin(\alpha_t) \approx W\alpha_t$, where α_t is the transversal inclination of the road, while the camber thrust is equal to the weight multiplied by the camber stiffness coefficient and the angle. It is clear that if the value of the camber stiffness coefficient is larger than one (measured in rad^{-1}), as is true of bias ply tires, the net force is directed uphill; the opposite occurs for radial tires. This situation occurs when a rut is present in the road: A radial tire tends to track in the bottom of the rut, while a bias-ply tire tends to climb out of it.

To include camber thrust into the linearized model, Eq. (2.27) can be modified as:

$$F_y = -C\alpha + C_\gamma \gamma \; . \tag{2.28}$$

The equation can be used with confidence for values of α up to about 4 deg and of γ up to 10 deg.

The self aligning moment can be expressed by a linear law:

$$M_z = (M_z)_{,\alpha}\alpha \; , \tag{2.29}$$

where $(M_z)_\alpha$ is the derivative $\partial M_z/\partial\alpha$ computed for vanishingly small α and γ and is defined as the aligning stiffness coefficient or simply the *aligning coefficient*.

Here again the effect of the camber angle can be included by modifying Eq. (2.29) as:

$$M_z = (M_z)_{,\alpha}\alpha + (M_z)_{,\gamma}\gamma \; , \tag{2.30}$$

where $(M_z)_{,\gamma}$ is the derivative $\partial M_z/\partial\gamma$ computed for vanishingly small α and γ, but the second effect is so small that it is usually neglected.

Equation (2.30) supplies a sound approximation of the aligning torque for a range of α far more limited than that in which Eq. (2.27) holds. Note however,

that the importance of the aligning torque in the study of the behavior of the vehicle is limited. Consequently, a precision lower than that required for side forces can be accepted. In practical terms, a good approximation of the aligning torque is important only when studying the steering mechanism.

The aligning stiffness coefficient due to sideslip angle is about 0.01 m/deg (Nm/N deg) for bias ply tires and 0.013 m/deg for radial tires, while that due to camber (aligning camber stiffness coefficient) is approximately 0.001 m/deg for the first and 0.0003 m/deg for the latter.

A small aligning moment is caused by the curvature of the path even when the sideslip angle is equal to zero; however, this effect is felt only if the radius of the trajectory is very small, on the order of a few meters, and consequently it is present only in low speed manoeuvres. It can be important for designing the steering system for the mentioned conditions.

The definition of the cornering coefficient implies that the cornering stiffness is linear with the normal load F_z; actually the cornering stiffness behaves in this way only for low values of force F_z and then increases to a lesser extent (Fig. 2.57). When the limit value has been reached it remains constant or slightly decreases. It is often expedient to approximate the cornering stiffness as a function of load with two straight lines, the second of which is horizontal. Note that in the figure the line corresponding to a sideslip angle of 2 deg refers to the true cornering stiffness, while the other curve ($\alpha = 10$ deg) is related to a sort of average stiffness.

When the need for a more detailed numerical description of the lateral behavior of a tire arises, there is no difficulty, at least in theory, in approximating the experimental law $F_y(\alpha, \gamma, F_z, p, V, ...)$ and the similar relationship for the aligning torque, using algorithms common to numerical analysis. This approach can be used with success in the numerical simulations of the behavior of the

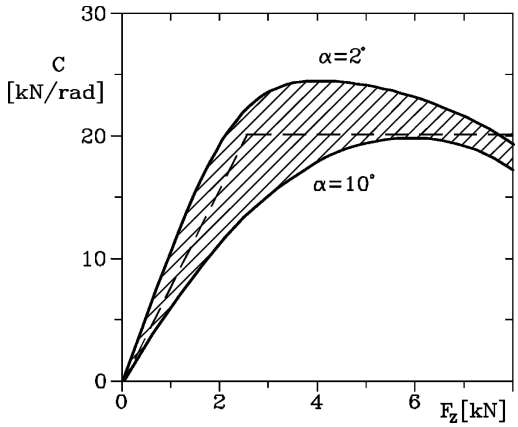


FIGURE 2.57. Cornering stiffness in function of load F_z (curve shown as $\alpha = 10$ deg is a sort of "secant" stiffness).

vehicle, even if it is often quite expensive in terms of the time needed for data preparation and computation. A problem common to many numerical approaches of this kind is that of requiring a great amount of experimental data, which are often difficult, or costly, to obtain.

Polynomial approximations, with terms including the third power of the slip angle α, can be used.

As previously stated, Eq. (2.25) can also be used to express the cornering force and the aligning moment as a function of the various parameters.

In the case of the side force, the magic formula is:

$$F_y = D \sin\left(C \arctan\left\{B(1-E)(\alpha+S_h)+E \arctan\left[B(\alpha+S_h)\right]\right\}\right)+S_v \ , \quad (2.31)$$

where the product of coefficients B, C and D yields the cornering stiffness directly. The values of the other coefficients are:

$$C = a_0 \qquad D = \mu_{y_p} F_z \ ,$$

where a value of 1.30 is suggested for a_0 and $\mu_{y_p} = a_1 F_z + a_2$,

$$E = a_6 F_z + a_7 \ ,$$

$$BCD = a_3 \sin\left[2 \arctan\left(\frac{F_z}{a_4}\right)\right](1 - a_5|\gamma|) \ ,$$

$$S_h = a_8 \gamma + a_9 F_z + a_{10} \ ,$$

$$S_v = a_{11} \gamma F_z + a_{12} F_z + a_{13} \ .$$

Product BCD is the cornering stiffness changed in sign.

To obtain a better description of the camber thrust, the constant a_{11} is often substituted by the linear law:

$$a_{11} = a_{111} F_z + a_{112} \ .$$

Coefficients S_h and S_v account for ply steer and conicity forces.

Similarly, in the case of the aligning torque the formula is:

$$M_z = D \sin\left(C \arctan\left\{B(1-E)(\alpha+S_h)+E \arctan\left[B(\alpha+S_h)\right]\right\}\right)+S_v \ , \quad (2.32)$$

$$C = c_0 \ , \qquad D = c_1 F_z^2 + c_2 F_z \ ,$$

where a value of 2.40 is suggested for c_0,

$$E = (c_7 F_z^2 + c_8 F_z + c_9)(1 - c_{10}|\gamma|) \ ,$$

$$BCD = (c_3 F_z^2 + c_4 F_z)(1 - c_6|\gamma|) e^{-c_5 F_z} \ ,$$

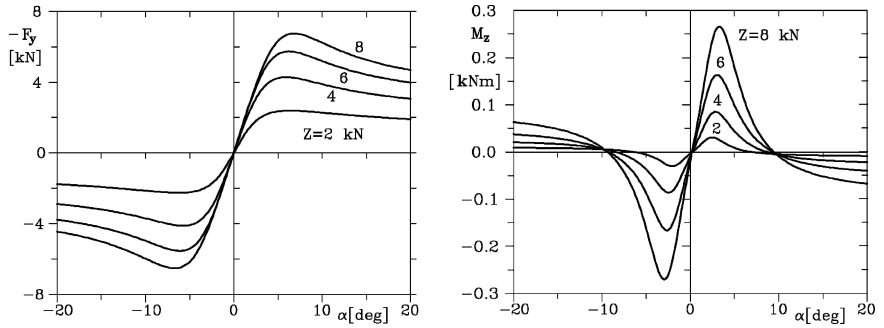


FIGURE 2.58. Curves $F_y(\alpha)$ e $M_z(\alpha)$ obtained by using the magic formula (2.31) and (2.32). Radial tire 205/60 R 15 V.

$$S_h = c_{11}\gamma + c_{12}F_z + c_{13} \ ,$$

$$S_v = (c_{14}F_z^2 + c_{15}F_z)\gamma + c_{16}F_z + c_{17} \ .$$

The units introduced into the magic formula (2.31) and (2.32) are usually not consistent: The load F_z is expressed in kN, angles α and γ are in degrees, F_y and M_z are obtained in N and Nm respectively.

The curves $Y(\alpha)$ and $N_a(\alpha)$ for values of the vertical load F_z equal to 2, 4, 6 and 8 kN for a radial tire 205/60 VR 15 6J are shown in Fig. 2.58.

The magic formula can be used to model the behavior of the tire more completely than by expressing the direct relationship of the side force and aligning moment with the sideslip angle. As an example, the carpet plots of the side force, the side force coefficient, the aligning moment, the camber thrust and the camber moment are shown in Fig. 2.59, along with the Gough diagram for the type of tire shown in Fig. 2.58.

It is also possible to build structural models of the tire to express the forces it exerts by examining the deformations and stresses its structure is subjected to. Apart from complex numerical models, primarily based on the finite element method, which allow one to compute the required characteristics but are so complex that they are of little use in vehicle dynamics computations, it is possible to resort to simplified models, treating the tread band as a beam or as a string on elastic foundations.[8] These models allow one to obtain interesting results, particularly from a qualitative viewpoint, as they link the performance of the tire with its structural parameters, but their quantitative precision is usually lower than that of empirical models, particularly those based on the magic formula, which is increasingly becoming a standard in tire modelling.

[8]See, for instance, J. R. Ellis, *Vehicle Dynamics*, Business Books Ltd., London, 1969; G. Genta, *Meccanica dell'autoveicolo*, Levrotto & Bella, Torino, 1993.

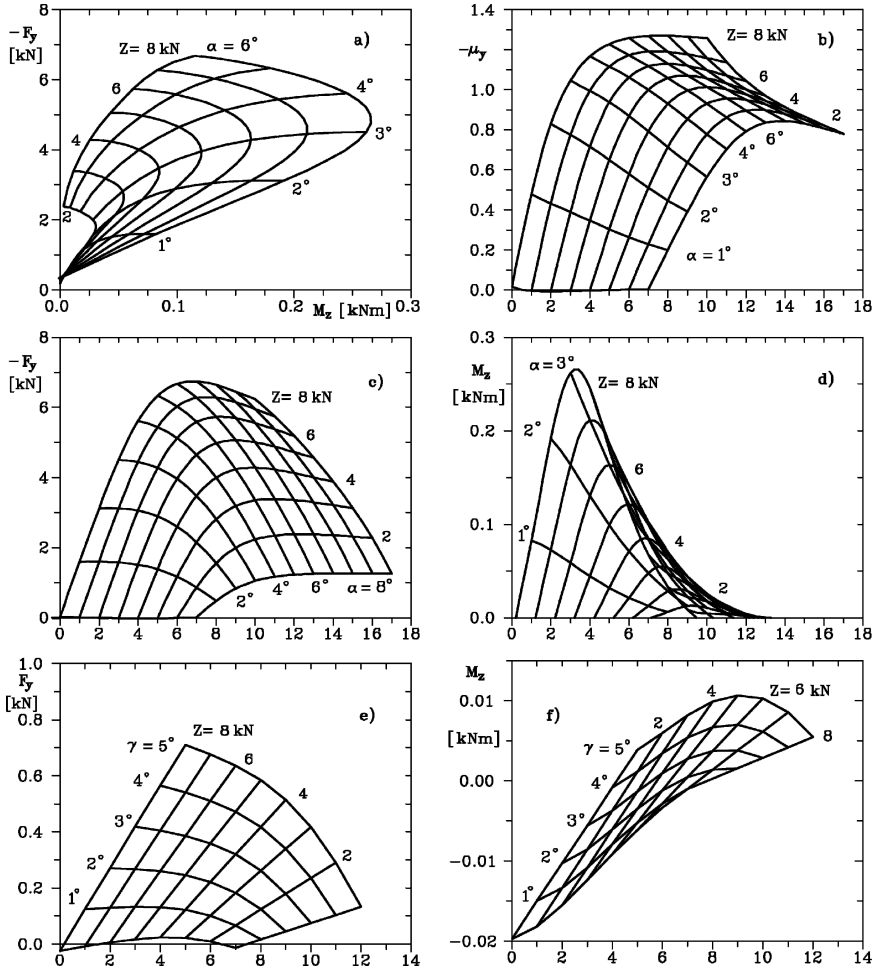


FIGURE 2.59. (a) Gough diagram and carpet plot (b) of sideslip coefficient, (c) cornering force, (d) self aligning torque, (e) camber force e (f) camber moment for a radial tire 205/60 R 15 V, obtained using the magic formula.

2.8 INTERACTION BETWEEN LONGITUDINAL AND SIDE FORCES

The considerations seen in the preceding sections apply only in the case in which longitudinal and side forces are generated separately. If the tire produces forces in the X' and Y' directions simultaneously the situation can change, as the traction used in one direction limits that available in the other.

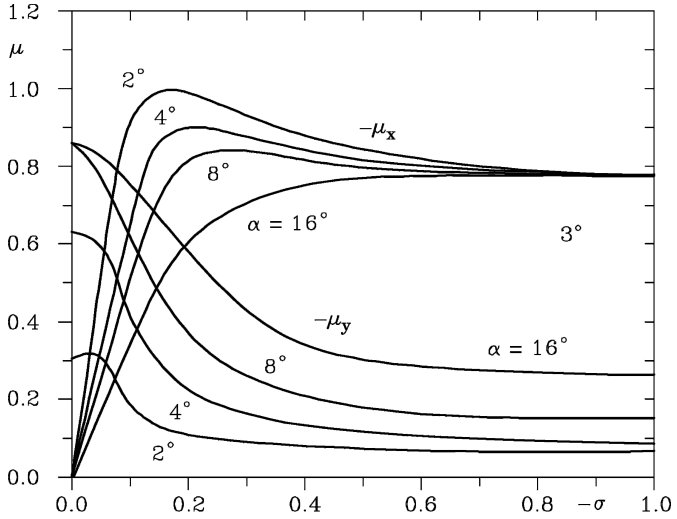


FIGURE 2.60. Lateral and longitudinal force coefficients as function of longitudinal slip and sideslip angle.

By applying a driving or braking force to a tire that has a certain sideslip angle, the cornering force reduces. The same applies to the longitudinal force a tire can exert if called to exert a lateral force as well.

An example is shown in Fig. 2.60: By setting the tire at a given sideslip angle and applying a braking torque, the shape of the curve $\mu_x(\sigma)$ is deeply changed. While the value in slipping conditions μ_s is almost unchanged, the peak value μ_p decreases to a greater extent.

It is then possible to obtain a polar diagram of the type shown in Fig. 2.61, in which the force in the Y' direction is plotted against the force in the X' direction for any given value of the sideslip angle α. Each point of the curves is characterized by a different value of the longitudinal slip σ. In a similar way it is possible to plot a curve $F_y(F_x)$ at constant σ.

Strictly speaking, the curves are not exactly symmetrical with respect to the F_y axis: Tires usually develop the maximum value of the force F_y when they exert a very light longitudinal braking force, and a slight braking force can actually increase the lateral force exerted at a moderate sideslip angle.

Together with the curves $F_y(F_x)$ two other curves $M_z(F_x)$ are also shown in Fig. 2.61. An application of driving forces results in an increase of the aligning torque, while a braking force causes the aligning moment to decrease to the point that it changes its sign when the braking limit is approached. This effect is destabilizing, as it tends to increase the sideslip angle.

A set of experimental curves $F_y(F_x)$ at constant α is shown in Fig. 2.62a.

If F is the total force exerted on the wheel by the road while F_x and F_y are its components, the resultant force coefficient can be expressed as:

$$\mu = \frac{F}{F_z} = \sqrt{\mu_x^2 + \mu_y^2}\,. \tag{2.33}$$

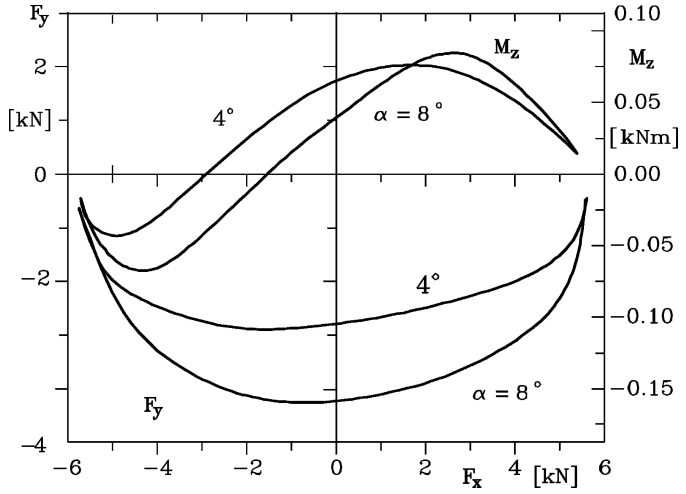


FIGURE 2.61. Polar diagrams of lateral force on the wheel at constant sideslip angle and self aligning torque as function of longitudinal force.

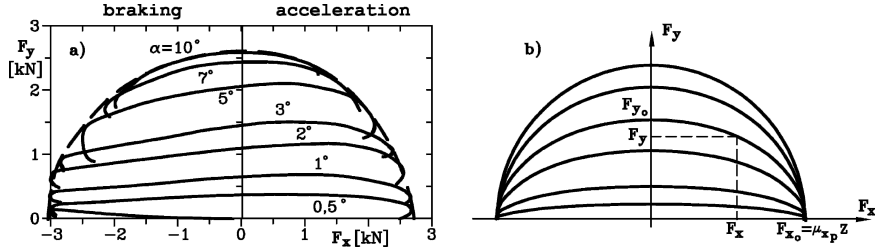


FIGURE 2.62. Polar diagrams of force at constant sideslip angle. (a) Experimental diagrams; (b) elliptic approximation.

The various curves plotted for different values of α are enveloped by the polar diagram of the maximum force the tire can exert. If it were a circle, the so-called *friction circle*, as in simple models it can be assumed to be, the maximum force coefficient would be independent of direction.

Actually, not only is the value of μ_x greater than that of μ_y but, as already stated, there is some difference in longitudinal direction between driving and braking conditions. The envelope, as well as the whole diagram, is a function of many parameters. Apart from the already mentioned dependence on the type of tire and road conditions, there is a strong reduction of the maximum value of force F with the speed, one that is particularly strong on wet road.

The three dimensional diagram of Fig. 2.63, limited to driving force F_x, gives an idea of this reduction. The phenomenon is enhanced by water, as in diagrams of Fig. 2.64.

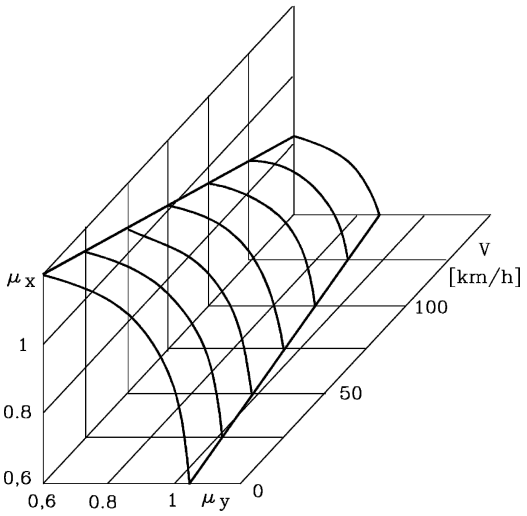


FIGURE 2.63. Variation of the envelope curve of polar diagrams of force between tire and road at different speeds.

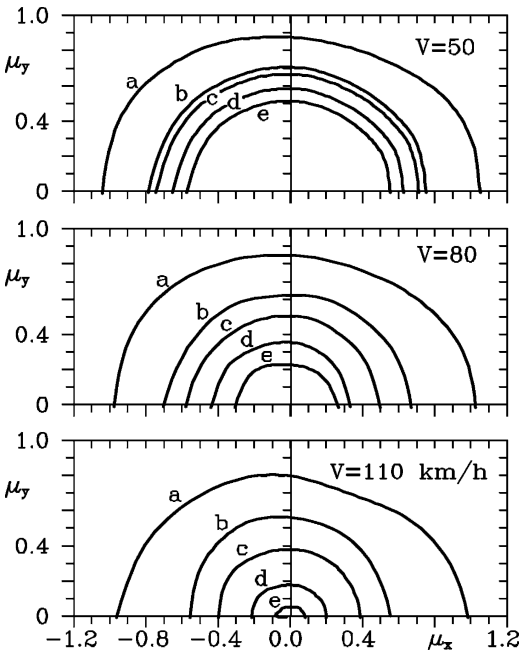


FIGURE 2.64. Envelope curve of $F_y(F_x)$ curves, at different road conditions. (a) Dry road and (b) wet with 0.2 mm; (c) 0.5 mm; (d) 1 mm and (e) 2 mm of water.

A model allowing us to approximate the curves $F_y(F_x)$ at constant α with simple functions can be quite useful. This can be obtained by using the elliptical approximation (Fig. 2.62b):

$$\left(\frac{F_y}{F_{y_0}}\right)^2 + \left(\frac{F_x}{F_{x_0}}\right)^2 = 1\,, \tag{2.34}$$

where forces F_{y_0} and F_{x_0} are, respectively, the force F_y exerted, at the given sideslip angle, when no force F_x is exerted, and the maximum longitudinal force exerted at zero sideslip angle. The envelope curve is then elliptical, the friction ellipse.

If Eq. (2.34) is used to express function $F_y(F_x)$, the cornering stiffness of a tire which is exerting a longitudinal force F_x can be expressed as a function of the cornering stiffness C_0 (i.e. the cornering stiffness when no longitudinal force is produced) by the expression:

$$\left(\frac{C\alpha}{C_0\alpha}\right)^2 + \left(\frac{F_x}{F_{x_0}}\right)^2 = 1\,. \tag{2.35}$$

The cornering stiffness can be expressed by the function:

$$C = C_0\sqrt{1-(F_x/\mu_p F_z)^2}\,, \tag{2.36}$$

where force F_{x_0} has been substituted by $\mu_p F_z$.

Although a rough approximation, particularly in cases where the longitudinal force approaches its maximum value (the differences between the curves of Fig. 2.62a and those of Fig. 2.62b are evident), the elliptical approximation is often used for all cases in which the concept of cornering stiffness is useful.

The empirical model of Eqs. (2.25) and (2.31) has been modified by Pacejka, to allow for the interaction between longitudinal and lateral forces in a better way than that of computing the two forces separately and then using the elliptic approximation.

First of all, the formulae for longitudinal force calculation, in absence of the sideslip angle, have been redefined; they refer, therefore, to pure traction or braking:

$$\begin{aligned} F_x &= F_{x0}(\sigma, F_z), \\ F_{x0} &= D_x \sin\Theta_x + S_{Vx}, \\ \Theta_x &= C_x \arctan\left\{B_x\sigma_x - E_x\left(B_x\sigma_x - \arctan\left(B_x\sigma_x\right)\right)\right\}, \\ \sigma_x &= \sigma + S_{Hx}, \\ \gamma_x &= \gamma\lambda_{\gamma x} \end{aligned} \tag{2.37}$$

Coefficients of previous formulae are calculated according to the following expressions:

$$\begin{aligned}
C_x &= p_{Cx1}\lambda_{Cx}, \\
D_x &= \mu_x F_z, \\
\mu_x &= (p_{Dx1} + p_{Dx2} df_z)\left\{1 - p_{Dx3}\gamma_x^2\right\}\lambda_{\mu x}, \\
E_x &= \left(p_{Ex1} + p_{Ex2} df_z + p_{Ex3} df_z^2\right)\left[1 - p_{Ex4}\frac{\sigma_x}{|\sigma_x|}\right]\lambda_{Ex}\ , \\
B_x C_x D_x &= F_z\left(p_{Kx1} + p_{Kx2} df_z\right)\lambda_{Kx} e^{p_{Kx3} df_z}, \\
S_{Hx} &= (p_{Hx1} + p_{Hx2} df_z)\lambda_{Hx}, \\
S_{Vx} &= F_z(p_{Vx1} + p_{Vx2} df_z)\lambda_{Vx}\lambda_{\mu x}.
\end{aligned} \tag{2.38}$$

Formulae are similar to the previous Eq. (2.25), except for the difference in coefficients p and the introduction of scale coefficient λ, which is in most cases set to 1.

Parameter df_z is given by the following equation:

$$df_z = \frac{F_z - F_{z0}}{F_{z0}},$$

where F_{z0} is a reference value for the vertical load.

The cornering force calculation formulae have been modified, to be used the absence of longitudinal slip (pure side slip):

$$\begin{aligned}
F_y &= F_{y0}(\alpha, \gamma, F_z), \\
F_{y0} &= D_y \sin\Theta_y + S_{Vy}, \\
\Theta_y &= C_y \arctan\left\{B_y\alpha_y - E_y\left(B_y\alpha_y - \arctan\left(B_y\alpha_y\right)\right)\right\}, \\
\alpha_y &= \alpha + S_{Hy}, \\
\gamma_y &= \gamma\lambda_{\gamma y}
\end{aligned} \tag{2.39}$$

Similarly, coefficients of previous equations are given by:

$$\begin{aligned}
C_y &= p_{Cy1}\lambda_{Cy}, \\
D_y &= \mu_y F_z, \\
\mu_y &= (p_{Dy1} + p_{Dy2} df_z)\left\{1 - p_{Dy3}\gamma_y^2\right\}\lambda_{\mu y}, \\
E_y &= (p_{Ey1} + p_{Ey2} df_z)\left\{1 - (p_{y3} + p_{Ey4}\gamma_y)\frac{\alpha_y}{|\alpha_y|}\right\}\lambda_{Ey}, \\
B_y C_y D_y &= p_{Ky1} F_{z0} \sin\left[2\arctan\left\{F_z/(p_{Ky2} F_{z0}\lambda_{Fz0})\right\}\right] \cdot \\
&\quad \cdot(1 - p_{Ky3}\left|\gamma_y\right|)\lambda_{Fz0}\lambda_{Ky}, \\
S_{Hy} &= (p_{Hy1} + p_{Hy2} df_z)\lambda_{Hy} + p_{Hy3}\gamma_y, \\
S_{Vy} &= F_z[(p_{Vy1} + p_{Vy2} df_z)\lambda_{Vy}] + \\
&\quad + F_z[(p_{Vy3} + p_{Vy4} df_z)\gamma_y\lambda_{\mu y}].
\end{aligned} \tag{2.40}$$

The formulae here are similar to the previous (2.31), with the exception of coefficients p and scale factors λ, which, again, are seldom other than 1. The parameter df_z is as previously given.

In the simultaneous presence of both forces the longitudinal force has to be modified by multiplying the result of Eq. (2.37) and (2.39) by the correction factor:

$$F_x = F_{x0} G_{x\alpha}(\alpha, \sigma, F_z)$$

The calculation of the correction factor is made by:

$$\begin{aligned}
B_{x\alpha} &= r_{Bx1} \cos\left\{\arctan\left(r_{Bx2}\sigma\right)\right\} \lambda_{x\alpha}, \\
C_{x\alpha} &= r_{Cx1}, \\
D_{x\alpha} &= \frac{F_{x0}}{\cos \Phi_x}, \\
E_{x\alpha} &= r_{Ex1} + r_{Ex2} df_z, \\
S_{Hx\alpha} &= r_{Hx1}, \\
G_{x\alpha} &= \frac{\cos \Psi_x}{\cos \Phi_x}, \\
\Phi_x &= C_{x\alpha} \arctan\left\{B_{x\alpha} S_{Hx\alpha} - E_{x\alpha}(B_{x\alpha} S_{Hx\alpha} \arctan(B_{x\alpha} S_{Hx\alpha}))\right\}, \\
\Psi_x &= C_{x\alpha} \arctan\left\{B_{x\alpha} \alpha_s - E_{x\alpha}\left(B_{x\alpha}\alpha_s - \arctan(B_{x\alpha}\alpha_s)\right)\right\}.
\end{aligned} \tag{2.41}$$

In this case we again have the coefficients r and the scale factor λ.
In the same way:

$$F_y = F_{y0} G_{y\sigma}(\alpha, \sigma, \gamma, F_z) + S_{Vy\sigma}.$$

The correction factor is given by:

$$\begin{aligned}
B_{y\sigma} &= r_{By1} \cos\left\{\arctan\left(r_{By2}(\alpha - r_{By3})\right)\right\} \lambda_{x\alpha}, \\
C_{y\sigma} &= r_{Cy1}, \\
D_{y\sigma} &= \frac{F_{y0}}{\cos \Phi_y}, \\
E_{y\sigma} &= r_{Ey1} + r_{Ey2} df_z, \\
S_{Hy\sigma} &= r_{Hy1} + r_{Hy2} df_z, \\
S_{Vy\sigma} &= D_{Vy\sigma} \sin\left[r_{Vy5} \arctan(r_{Vy6}\sigma)\right] \lambda_{Vy\sigma}, \\
D_{Vy\sigma} &= \mu_y F_z (r_{Vy1} + r_{Vy2} df_z + r_{Vy3}\gamma) \cos\left[\arctan(r_{Vy4}\alpha)\right] \\
G_{x\alpha} &= \frac{\cos \Psi_y}{\cos \Phi_y}, \\
\Phi_y &= C_{y\sigma} \arctan\left\{B_{y\sigma} S_{Hy\sigma} - E_{y\sigma}(B_{y\sigma} S_{Hy\sigma} \arctan(B_{y\sigma} S_{Hy\sigma}))\right\}, \\
\Psi_y &= C_{y\sigma} \arctan\left\{B_{y\sigma}\sigma_s - E_{y\sigma}\left(B_{y\sigma}\sigma_s - \arctan(B_{y\sigma}\sigma_s)\right)\right\}.
\end{aligned} \tag{2.42}$$

This method of modelling the interaction between F_x and F_y is quite complicated and not well justified by theory.

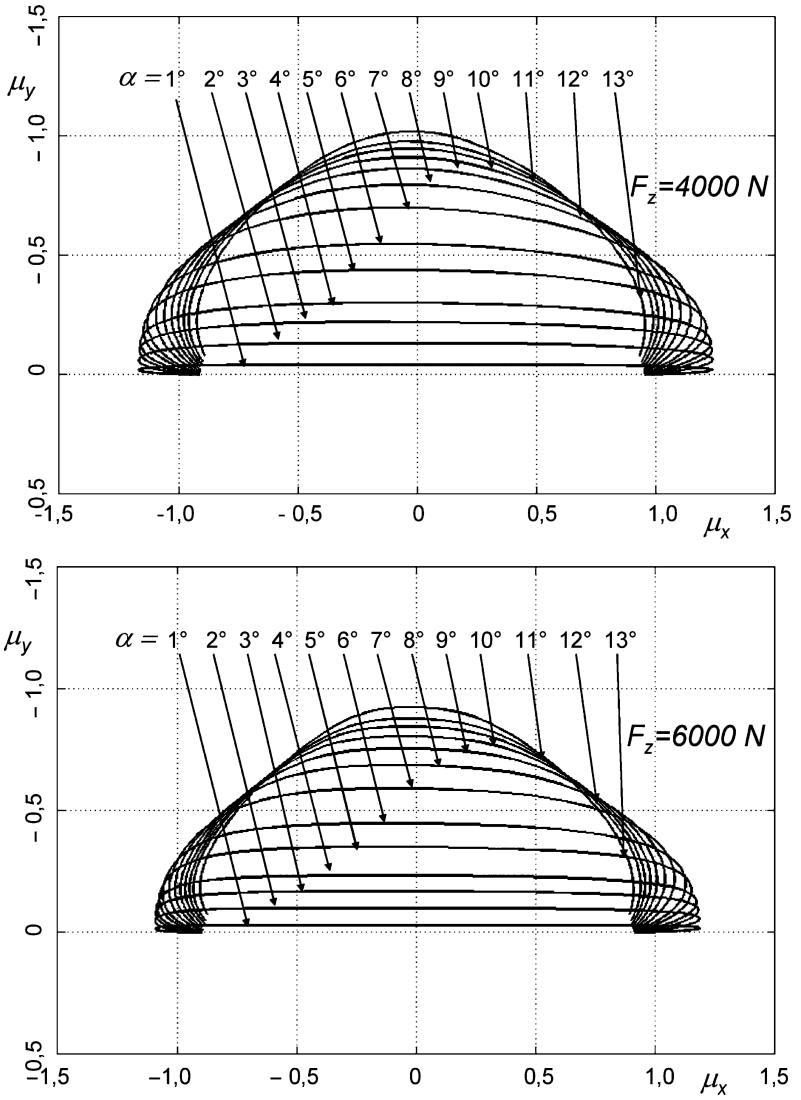


FIGURE 2.65. Ellipses calculated with modified Pacejika formulae for the same 195/65 R 15, tire, used to calculate carpet plots previously shown.

Results are reasonable, better than those obtained by the elliptic approximation only; nevertheless there are cases where curves $F_y(F_x)$ have non-realistic shapes and the above formulae must be carefully applied.

Research activities in this field are ongoing, with the goal of obtaining a simpler and more realistic set of formulae.

A graph similar to that of Fig. 2.62 and calculated with this method is shown in Fig. 2.65.

2.9 OUTLINE ON DYNAMIC BEHAVIOR

2.9.1 Vibration modes

The dynamic behavior of the tire is quite important in determining both the comfort and stability of the vehicle. Although the strong interactions between the vibrational behavior of the tires and suspensions and the suspended mass of the vehicle suggest that the dynamic study should be performed on the complete vehicle, it is nonetheless interesting to study the behavior of the tire alone, if only to obtain the data which will later be introduced in more complex models.

As the stiffness of the tire is greater than that of the suspension, the behavior of the former is generally not significant in low frequency motions ($1 \div 3$ Hz), for which the tire can be modelled as a rigid body.

In an intermediate field of frequencies ($10 \div 20$ Hz) the tire can be considered as a massless deformable element introduced into the system. It is possible to define a dynamic stiffness in both vertical and transversal direction. It is possible to show that the tire should be characterized by a low vertical stiffness, in order to minimize vertical displacements of the suspended mass on irregular road surfaces, and high transversal stiffness, to react with small displacements to side forces applied to the vehicle.

It is evident that there is a notable interaction between the vertical behavior of the tire and that of the rest of the vehicle, between transversal stiffness and cornering stiffness and, finally, between vertical and lateral stiffness, because of the elasto-kinematic behavior of the suspension.

In Fig. 2.66 the dynamic vertical and lateral stiffness of two 155 15 tires are shown as a function of speed. As can be seen, the dynamic stiffness of radial tires is less than that of bias ply tires.

At higher frequencies (over 50 Hz) the tire vibrates according to its natural frequencies. In Fig. 2.67 deformation profiles are shown corresponding to forced vibrations of 155 15 tires of both radial and cross ply type. Angles ϕ_r, shown in the picture, are the phase retards between tire and exciting platform vibration. Radial tires present a peak between $60 \div 90$ Hz, depending upon tire dimension, while other less relevant peaks are present at higher frequencies, with a reduced amplification factor.

Transmissibility, defined as the ratio between response and excitation amplitudes for a radial and a cross ply tire, is shown in Fig. 2.68.

The figure shows the vertical response to a contact point vertical motion: In the same way it is possible to measure horizontal transmissibility.

The resonance of the unsprung mass caused by the tire's vertical elasticity takes place at a much lower frequency; the first resonance peak of the radial tire shown in the figure is bound to a mode where the tread moves like a rigid body on the sides. Higher frequency modes where the tread vibrates according to lobate shapes are less excited and more dampened.

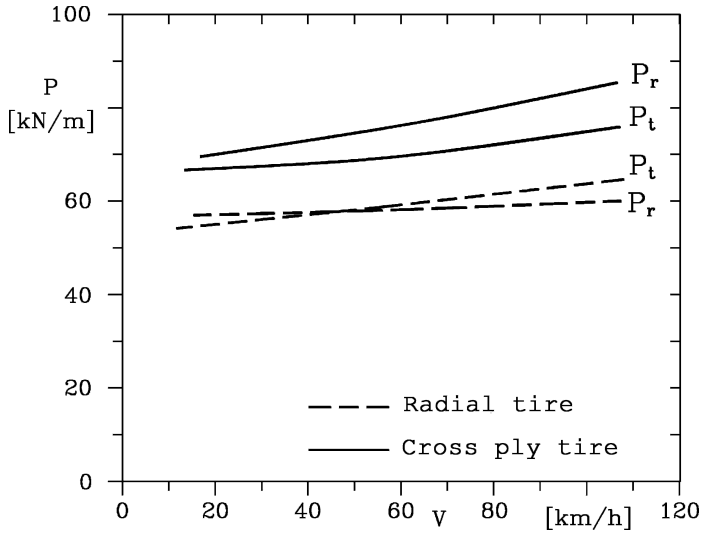


FIGURE 2.66. Dynamic stiffness in vertical P_r and lateral P_t direction of two 155 15 tires of radial and cross ply type, in function of speed.

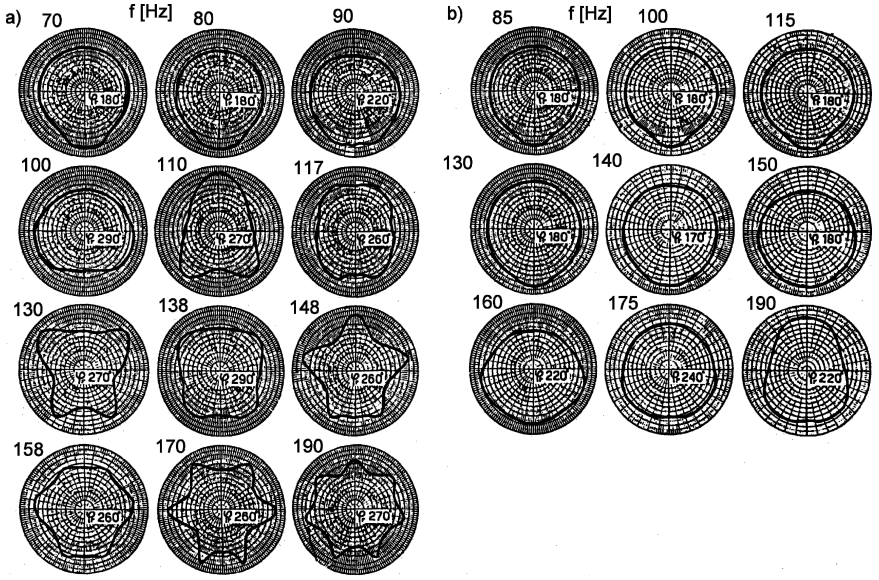


FIGURE 2.67. Deformation of a 155 15 radial tire (a) and same size cross ply tire (b) at different excitation frequencies.

2.9.2 Cornering dynamic forces

If the geometrical parameters (slip and camber angles) or the forces in the X' and Z' directions are variable during motion, the values of the side force and

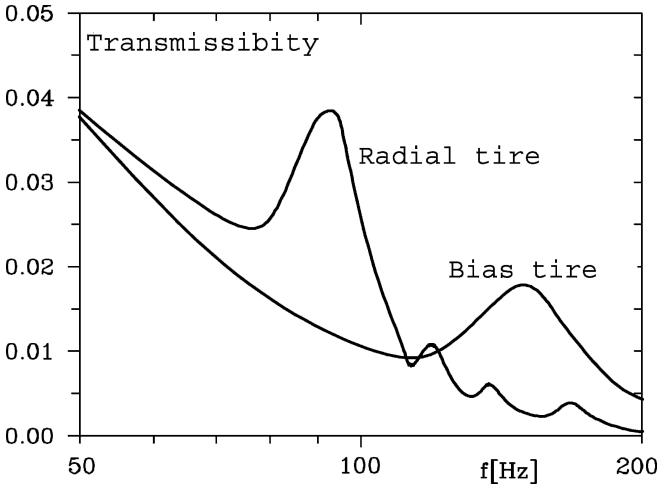


FIGURE 2.68. Transmissibility for a radial and a bias tire 155 R 15 and 155 15 as a function of frequency. A frequency interval between 50 and 200 Hz is shown; the resonance of the unsprung mass for tire elasticity takes place at a lower frequency.

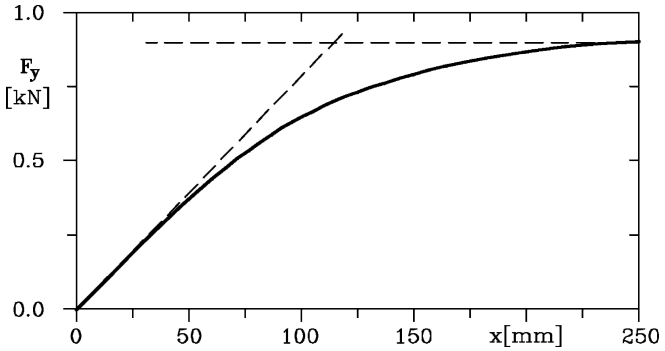


FIGURE 2.69. Lateral forces applied to a standing tire then put in rotation for various times or distances. Tire 140 12, $F_z = 3.5$ kN, $p = 138$ kPa.

aligning moment are at any instant different from those that would characterize a stationary condition with the same values of all parameters. As an example, if a tire is tilted about the vertical axis at standstill and then allowed to roll, the side force reaches the steady-state value only after a certain time, after rolling for a certain distance (Fig. 2.69) usually referred to as the *relaxation length*.

This effect is usually not noticeable in normal driving as the time delay is small, but the fact that there is a delay between the setting of the sideslip angle and the force generation is important in dynamic conditions.

If the sideslip angle is changed according to a harmonic law in time, the side force and aligning torque follow the sideslip angle with a certain delay, a

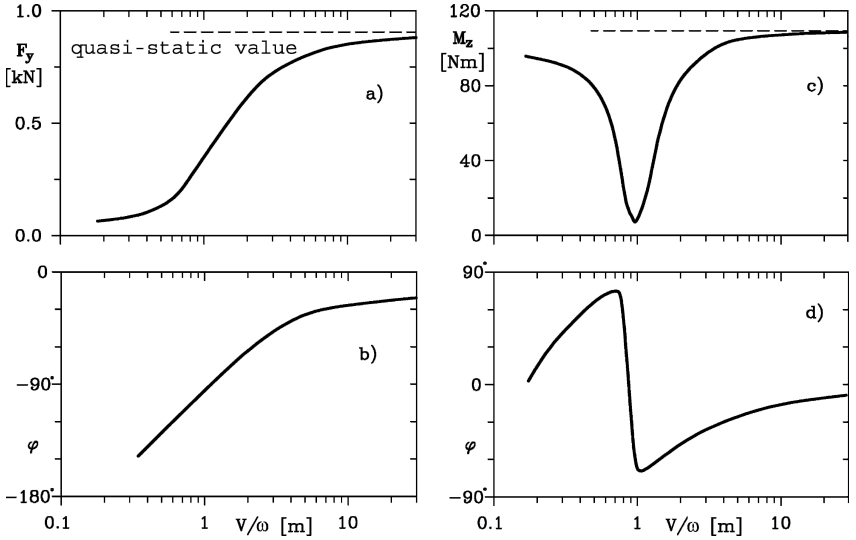


FIGURE 2.70. Lateral force and self aligning torque on a tire whose sideslip angle is changing according to harmonic law, between -4 deg and 4 deg, as a function of the ratio V/ω between speed and frequency of the low $\alpha(t)$, the wave length. Peak value (a), (c) and phase difference (b), (d) of force and moment are reported. Tire 190 14, $F_z = 4.8$ kN, $p = 165$ kPa.

function of the frequency, and their value is lower than that obtained in quasi-static conditions, i.e. at a very low frequency.

If the frequency is not high, at the speeds encountered in normal driving, the average values are not significantly lower than those characterizing static conditions, but a certain phase lag between the sideslip angle and the F_y force remains (Fig. 2.70).

The plots reported in the figure have been obtained for a given tire and are not easily generalized, particularly at high frequency. If a resonance of the tire occurs (as the phase of the aligning torque seems to suggest in the plot), the response is no longer dependent on the ratio V/ω but on the frequency and, to a lesser extent, on the speed.

More important in terms of practical applications is the case in which the load F_z applied by the wheel on the ground is variable, as is the case of rolling on uneven road (Fig. 2.71). The frequency can be high, if the speed is high enough, and the decrease of lateral force due to dynamic effects can be large. In the figure the law $z(t)$ of the vertical displacement of the hub of the wheel is harmonic, with a frequency of about 7 Hz, while the response $F_y(t)$ is more complicated. An inversion of sign can even occur at each cycle. The decrease of the average value of the lateral force with increasing frequency is shown in Fig. 2.71b.

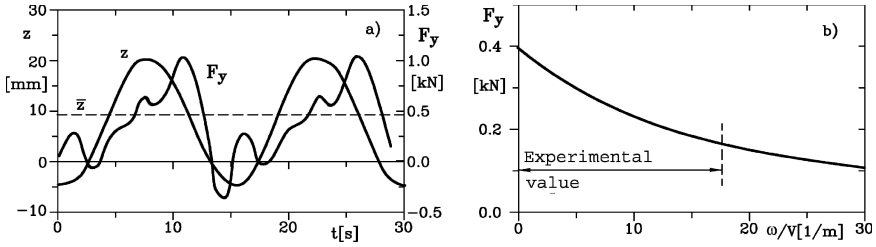


FIGURE 2.71. (a) Lateral force on a tire working with constant side slip angle, but with a rim moving vertically according to harmonic law $z(t)$. (b) Mean value of the force F_y as a function of the ratio between ω (referred to $z(t)$) and V.

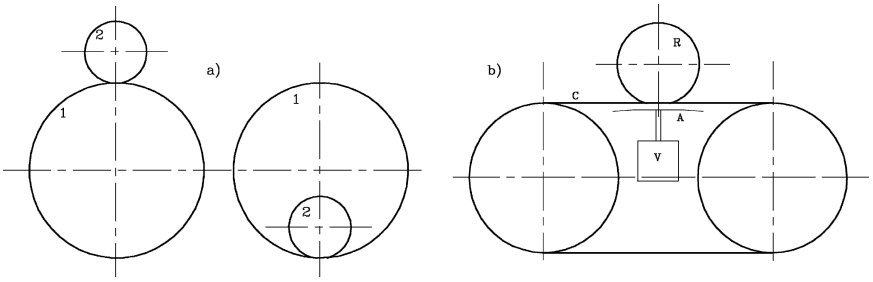


FIGURE 2.72. Tire testing machines. (a) Drum machines with positive and negative drum curvature. 1: drum, 2: tire to be tested. (b) Belt machine. A: belt bearing, C: belt, R: tire to be tested, V: excitation.

2.10 TESTING

The characteristics of tires can be measured using road or laboratory tests. The most common testing machines for tires are based on a drum that simulates the road; the tire can roll on the outer or inner surface of the drum (Fig. 2.72a).

The actual conditions are in a way intermediate between those encountered on the two types of test machines. To avoid the differences between the contact condition occurring in drum machines and actual use, more modern machines use a steel belt. The belt is kept flat by a hydrostatic bearing, which can be connected to a shaker to simulate motion on uneven road. The simulation of the road surface is easier on drum than on belt machines, whose surface is usually plain metal.

Other machines use a flat disc: Here the contact surface is flat but the tire works with a ground moving along a circular path, generating a side force even with no sideslip. For low speed and low duration tests the wheel can also roll against a moving platform. With platforms it is possible to simulate different road surfaces easily (Fig. 2.73).

For low speed tests of short duration the wheel can roll on a flat moving platform; different road surfaces can, in this case, be easily simulated.

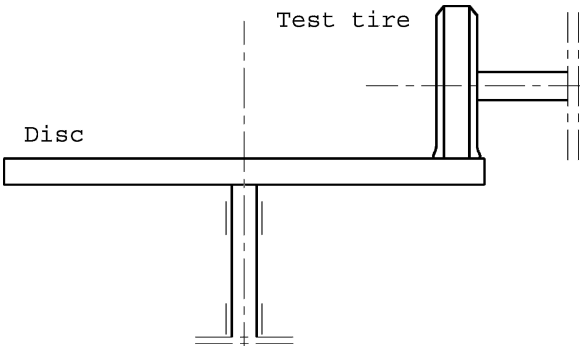


FIGURE 2.73. Testing machine where road surface is simulated through a flat disc.

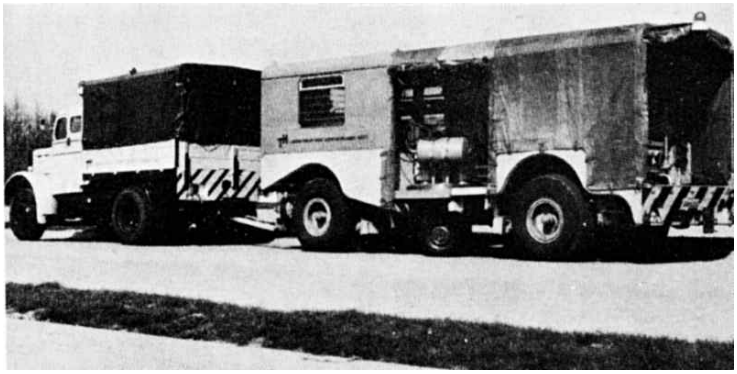


FIGURE 2.74. Vehicle for on road tire testing used by the vehicle research laboratory of the University of Delft. The test wheel is in the middle of the trailer.

In all test machines the tire is fit onto a hub able to measure the three force and three moment components. Different driving and braking conditions can be simulated using two different motors to move the wheel and the road. Wheels can be maintained at given camber and sideslip angles.

Laboratory tests, particularly those with roller machines, must be compared with those made on the road by dedicated test vehicles.

To test tires on the road, a wheel mounted on a dynamometric hub and provided with an independent motor is installed on a vehicle or a trailer. Figure 2.74 shows a picture of such a vehicle, used by the laboratory for vehicle testing at the University of Delft.

Such vehicles feature all the necessary equipment for tire force and moment measurement in every possible condition. An identical wheel is installed with an equal but opposite sideslip angle, in order to allow the vehicle to drive in a straight line. A suitable device can wet the road according to a determined water layer thickness.

To measure the rolling resistance only simpler machines can be used.

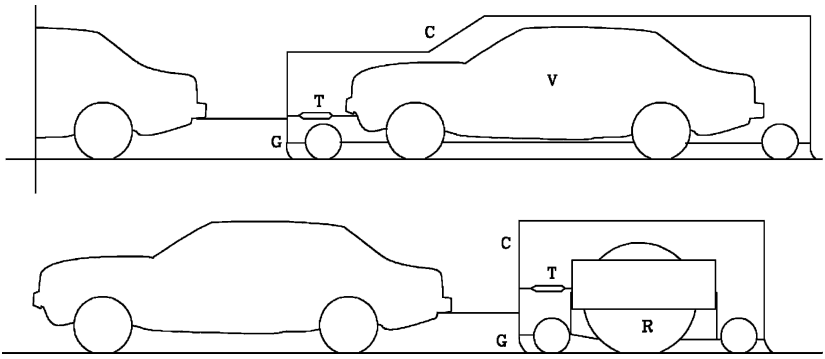


FIGURE 2.75. Fixture to measure rolling resistance of a vehicle or a wheel on the road. C: trailer, G: trailer skirt, R: wheel under test, T: dynamometric pull bar, V: vehicle under test.

On a wheel drum machine the wheel can be accelerated to a given speed, and then left to slow by its own inertia.

The equation for the motion of the wheel and drum assembly is:

$$\frac{d\Omega_R}{dt}\,\frac{J_R + J_S}{R_e} = -fF_z - \frac{|M_S|}{R_S}\,, \tag{2.43}$$

where R_S, J_S and M_S are the drum radius, its moment of inertia, reduced to the wheel axis, and the total resistant moment of the wheel. M_S can be determined the deceleration of the wheel, under the vertical force F_z, which can be obtained with a simple vertical load.

The $\Omega_R(t)$ law can be recorded easily and the roll resistance f calculated using the Eq. (2.43). This method has the inconvenience, common to all methods starting from the acquisition of a coast down curve, of requiring the measurement of a magnitude, where the derivative is desired. Because the derivative of a function is sensitive to errors of the function measurement, this method requires particular precautions.

Other inconveniences are common to all tests on the drum, particularly differences in the surface, usually of steel, and the shape of the contact patch.

Moreover, as in all coast down tests, the measurement is made during a transient, where many magnitudes are changing during the test, particularly the temperature.

If the acceleration phase is sufficiently long, the tire reaches the steady state temperature at a speed lower than that at test start. If the temperature is higher than the appropriate, rolling resistance is slightly lower. If the acceleration is higher, on the other hand, tire temperature may be too low, leading to a higher value of rolling resistance.

On road tests can be performed pulling a vehicle or a road, enclosed in a suitable trailer configured to eliminate aerodynamic resistance in its interior

(Fig. 2.75). A dynamometric pull bar connects the trailer with the test vehicle, allowing measurement of the rolling resistance.

This method has the inconvenience of being sensitive to road slope: A slope of only 0.1% can lead to errors of 10% in the rolling resistance coefficient f. The measurement can be made more difficult by possible longitudinal vibrations of the system.

3
SUSPENSIONS

3.1 INTRODUCTION

By vehicle suspension we mean a mechanism that links the wheel directly to the body or to a frame attached to the it.

Because a rigid vehicle with more than three wheels is a hyperstatic system, it is necessary for the vehicle structure to be flexible enough to allow the simultaneous contact of the wheels with the ground. Alternatively the wheels may be connected to a rigid body through a deformable system, the suspension. This second solution is adopted by most vehicles, while the first, widely found in horse carriages in the past, is now used on particularly slow vehicles.

In many cases, to suspension deformation must be added structural deformation, which plays an important role in the handling and comfort characteristics of a vehicle.

To accomplish their task, suspensions must:

- Allow a distribution of forces, exchanged by the wheels with the ground, complying with design specifications in every load condition
- Determine the vehicle *trim* under the action of static and quasi-static forces

We should not, in fact, forget that by introducing a deformable linkage on a vehicle, geometric variations of the body position are introduced as a function of payload and payload position; these variations are described through the three coordinates of the center of gravity and the three angles of the body reference system (yaw, roll, pitch angle). They are included under the term vehicle static

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

trim. A complete definition of the body reference system is given in the fourth part of this book, in volume two.

In addition to this function, which is accomplished basically with an elastic system, there is another, not less important function:

- To absorb and smooth out shocks that are received by the wheel from road irregularities and transmitted to the body

It should be remembered that this task requires the application of a suitable damping system; this function is so important that suspensions are also applied to two or three wheel vehicles that are not hyperstatic bodies.

In theory, tires alone could isolate the vehicle body from forces coming from the road, but their elastic and damping properties are not sufficient to achieve suitable handling and comfort targets, unless at very low speed and on smooth roads.

Suspensions are therefore essential for achieving adequate road holding behavior (*handling*) and comfort. They determine the distinctive characteristics of each vehicle.

Wheels must be, therefore, free to move in a direction almost perpendicular to the ground, in addition to their rotation and steering motion.

This vertical motion must be managed through the suspension linkages in order to guarantee the tire's correct position with reference to the ground. The capacity of a tire to react to suitable forces is, in fact, determined by the angles between the equator plane of the wheel with the ground and with the hub speed.

If the suspension is a kind of filter between road and body, designed to limit the amount of forces produced by road irregularities and maneuvers, this filter must not impair vehicle controllability in possible driving situations. Road holding depends not only on the mass properties of the vehicle (mass and moments of inertia), on its geometric properties (kind of traction, center of gravity position, wheelbase, track), and on its tires, but on suspensions, too.

Suspensions are usually divided into three classes: *independent*, *dependent* and *semi-independent* suspensions.

The first class presents no mechanical linkage between the two hubs of the same axle; a force acting on a single wheel does not affect the other; steering linkages, anti-roll bars or auxiliary frames are not taken into account in this description.

Dependent wheel suspensions or rigid axles provide for a rigid linkage between the two wheels of the same axle; each motion of a wheel caused by road irregularities affects the coupled wheel as well.

Semi-rigid suspensions have intermediate characteristics between the other two categories. In these suspensions wheel hubs cannot be considered independent because they are not linked with an articulated structure; indeed this structure's mechanical characteristics ensure that flexibility cannot be neglected. In practice, this category includes so-called *twist axles*.

Another important characteristic within these families separates steering from non-steering suspensions; while independent suspensions can be, in

principle, designed to become steering suspensions, dependent suspensions are no longer applied to steering axles, with the exception of industrial vehicles and off-road vehicles. The same applies without exception to semi-rigid suspensions.

Most suspensions can be applied to both driving and idling axles without impact on suspension design, other than for joint stiffness.

Considering elastic and damping systems only, suspensions can be classified as *passive* or *active*. In the first family the elastic reaction of the suspension system is determined solely by its deformation and the damping system can only waste part of the received energy. In the second family the suspension system can receive energy from other sources (the engine, or some intermediate storage of engine energy) to affect body motion, with the objective of limiting this motion close to its static equilibrium condition.

The objective of this kind of suspension is to limit body displacement to a minimum, necessary for transferring to the driver a sufficient perception of vehicle stability. We will comment on this kind of suspension in the chapter dedicated to chassis control systems and we will see that different levels of activity are identified, according to whether the outside energy source is used to correct the static trim only, or the dynamic trim as well. Most vehicle suspensions are passive.

Finally, two additional definitions.

We define as *sprung mass* that part of the vehicle mass that is free to move with reference to the ground, because of the suspension application. That part of the mass that does not change its position is called the *unsprung mass*.

Some suspension components contribute partly to sprung mass, partly to unsprung mass. To evaluate the two contributions, the mass of these elements must be divided into two parts, concentrated ideally in the suspension joints, in such a way as to conserve the moment of inertia and the center of gravity position.

3.1.1 Suspension components

To accomplish the functions described above, suspensions are built with different categories of components.

Bearing components or linkages

These are part of the mechanism linking the wheel to the body and guarantee the degrees of freedom of the wheels and their correct positions with reference to the ground. They determine the relative motion of the wheel with reference to the vehicle body; they also transfer to the body part of the load delivered from the tire contact patch.

Primary elastic members

These include springs (coil, bar and leaf springs), anti-roll bars and stop springs. These members connect the wheel to the body elastically and store the energy

produced by an uneven road profile. They not only store this energy but determine body position as a function of payload entity and position.

Secondary elastic members

These are elastic bushings on linkage joints. To some of these joints is given a certain elastic compliance.

This property was originally seen as a drawback to avoiding joint lubrication by using elastomeric joints.

More recently it was understood that it could be exploited to improve the design of the *elasto-kinematic* behavior of the suspension and its comfort properties.

The deformation of these joints plays an important role in determining vehicle handling.

Damping members

These are basically shock absorbers, but we should remember that primary and secondary elastic members also have a non-negligible capacity of wasting energy.

Shock absorbers are provided to waste the elastic energy stored by the elastic members and to allow the oscillation damping of the vehicle body, avoiding stationary vibration or resonance.

3.1.2 Suspension influence on body motion

Ideally, the suspensions should allow the wheels to move with respect to the body of the vehicle in a direction perpendicular to the road, maintaining the plane of the wheel parallel to itself and constraining all motions in the x and y directions[1]: A suspension of a single wheel should be a system with a single degree of freedom, the displacement in the z direction or suspension stroke.

It should be remembered that if the wheel is allowed to move along the Z' axis only, linkages should not transmit body rotation to the wheel hub, particularly roll angle.

None of the linkages we will examine, however, is able to satisfy this condition completely and every kind of car suspension has its own characteristics; the approximation to the ideal characteristic is part of the character of a car.

Wheel position is also affected by the stiffness of joints. Even spherical joints or cylindrical bushings are rather deformable; sometimes linkages too have to be considered as flexible members.

Wheel position is not only affected by body motion but also by external forces.

---

[1] Usually the reference system adopted for suspensions is the same as the reference system used on body drawings; this is the vehicle reference system, explained in the second volume, but the origin of this system is set in the symmetry plane of the body, though the front wheel centers in the static load position.

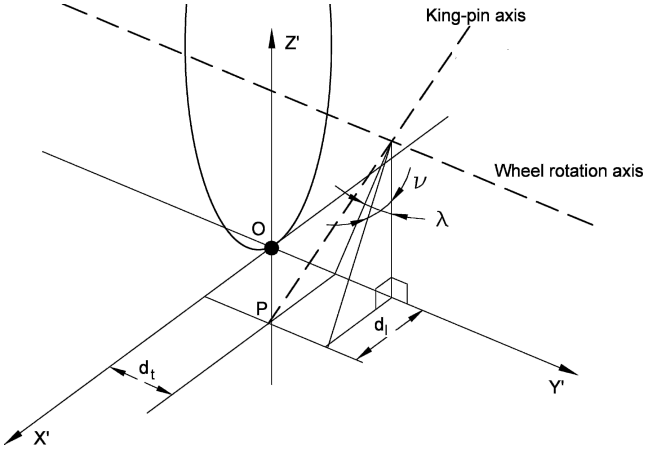


FIGURE 3.1. Steering geometry.

In the following lines we will define angles and linear dimensions applied to describe the behavior of suspensions; these characteristics are described with reference to a front steering axle.

In a steering suspension, the *steering axis* or *king-pin axis* is the axis around which the hub holder or strut, and therefore the wheel is able to rotate or steer (Fig. 3.1). Linkages of non-steering suspensions can leave to the wheel a limited freedom to steer around a king-pin axis, depending on body roll angle or applied forces; the effect of this additional steering angle will be discussed later.

The angle between the king-pin axis and the Z' axis, projected on the $X'Z'$ plane, is called the *caster* angle (ν in figure). As will be explained later, this angle is relevant to determining steering wheel return capabilities.

The distance between O, the center of the contact patch of the tire, and P, the intersection of the steering axis with the ground, projected on the $X'Z'$ plane is the *longitudinal trail* (d_l on figure).

The angle between the wheel equator plane and the $X'Z'$ plane is the inclination or *camber* angle, while the angle between the king-pin axis and the $Y'Z'$ plane is the *king-pin* angle (λ on figure).

We should remember that a suspension linking the wheel in such a way as to remain parallel to the car body (as, for example, in longitudinal trailing arms) applies to the wheel an inclination angle, with reference to the ground, equal to the body roll. This is particularly noxious, because the roll angle is such as to decrease the cornering stiffness of the tire. On two wheel vehicles, the behavior is opposite and advantageous; the same could be true on variable trim vehicles, which will be described in the second volume.

The distance between O and P projected on a $Y'Z'$ plane is the *king-pin offset* (d_t on figure).

But an ideal suspension should not allow the wheel to change its camber with reference to the ground; we will define as *camber recovery* angle the difference between the actual camber, with reference to the ground, and body roll. An ideal suspension should have a camber recovery equal to body roll.

With reference to the vehicle body, we define as *toe* angle the angle between the intersection of the equator plane of the wheel with the ground and the car x axis. It can be measured either in degrees or in millimeters of difference in transversal distance between two points on the wheel rim. It is called *toe-in* when the wheel equator planes cross in front of the vehicle and *toe-out* in the opposite case.

Other dimensions to be considered are the *wheelbase*, the distance projected on a xz plane between the center of tire contact area of the wheels of the same side and *track*, the distance projected on a yz plane between the centers of the contact area of the wheels of the same axle. Quite often, referring to suspension geometrical properties, *semi-wheelbase* and *semi-track* are distances of the center of the contact area from the xz and from the yz planes for the center of gravity of the vehicle.

When the vehicle body is moved in the vertical direction (displacement z) or rotates about its x axis (roll angle ϕ), the position of the wheel changes. It is possible to plot camber angle γ, track t, characteristic angles of the steering system, steering angle δ, etc. as functions of z and ϕ. These functions are generally strongly nonlinear but they can be linearized about any equilibrium position and the derivatives $\partial t/\partial z$, $\partial t/\partial \phi$, $\partial \gamma/\partial z$, $\partial \gamma/\partial \phi$, $\partial \delta/\partial z$, $\partial \delta/\partial \phi$, etc.[2] can be easily defined. They can be considered as constant in the small motions about an equilibrium position and define the behavior of the suspension.

When outlining the design of a new suspension, mathematical models are today fundamental.

With a simple approach, (see Fig. 3.2) beginning with suspension drawings, a suspension scheme can be built up, where linkages are represented with rigid beams connected with kinematic couples (spherical joints, bushings, etc.) able to simulate kinematic behavior by suspension stroke and wheel steering, taking into account the contribution of linkages and primary elastic members only.

Using a second approach, kinematic couples introduced into the model can be exchanged for elastic elements; it is now possible to calculate characteristic variations of the suspension as a function of applied forces as well. This second way of studying the suspension involves analyzing the elasto-kinematic behavior.

This simple simulation applied iteratively has a fundamental importance in designing suspensions with respect to assigned targets.

In the following paragraphs we will analyze the most widely used suspension schemes existing in present cars and their main components.

---

[2]In the following the notation $(t)_{,z}$, $(t)_{,\phi}$, etc. will be used.

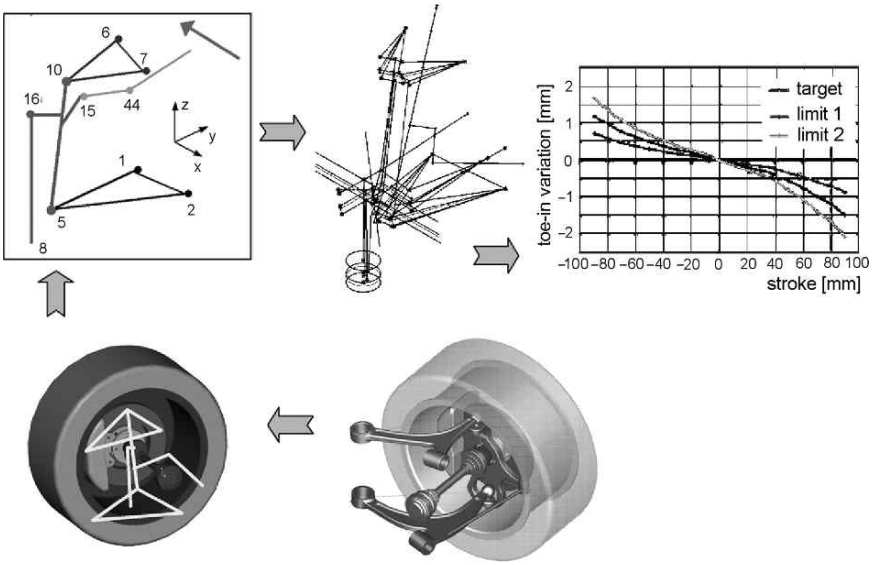


FIGURE 3.2. Beginning with suspension drawings a suspension scheme can be built up, where linkages are schematized with rigid beams connected with kinematic couples (spherical joints, bushings, etc.) able to simulate kinematic behavior by suspension stroke and wheel steering.

3.2 INDEPENDENT SUSPENSIONS

If the wheels are suspended independently, linkages must constrain five out of the six degrees of freedom of the wheel (or better, of the wheel hub, because the wheel is then free to rotate about its axis). The unconstrained degree of freedom should be the translation in a direction perpendicular to the ground. None of the many devices which are currently used fulfills this requirement exactly.

As suspensions must restrain five degrees of freedom, they can be materialized as a system made up of five bars with spherical hinges at the ends (Fig. 3.3), This layout, often referred to as multilink suspensions, has the advantage of allowing a large freedom of adjustment by changing the length of the bars by screwing the joints in or out. But it has little application outside the field of luxury cars because of its complexity, even if simpler multilink suspensions are now more widespread. From five bar multilink suspensions almost all configurations can be obtained by grouping these bars in different ways.

Note that in general the motion of the wheel is not planar and as a consequence the study of the kinematic behavior is not simple. It is, however, easy enough to obtain the exact kinematics of any suspension by using computer generated trajectories.

If points 1 and 2 and points 3 and 4 coincide, the corresponding bars become triangular elements: The suspension thus obtained is a transversal quadrilateral

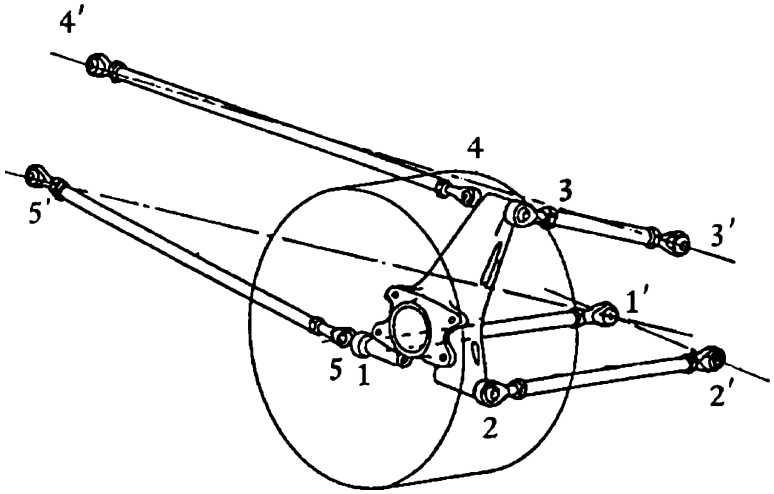


FIGURE 3.3. Suspension based upon five linkages which constrain five of the six degrees of freedom of the wheel.

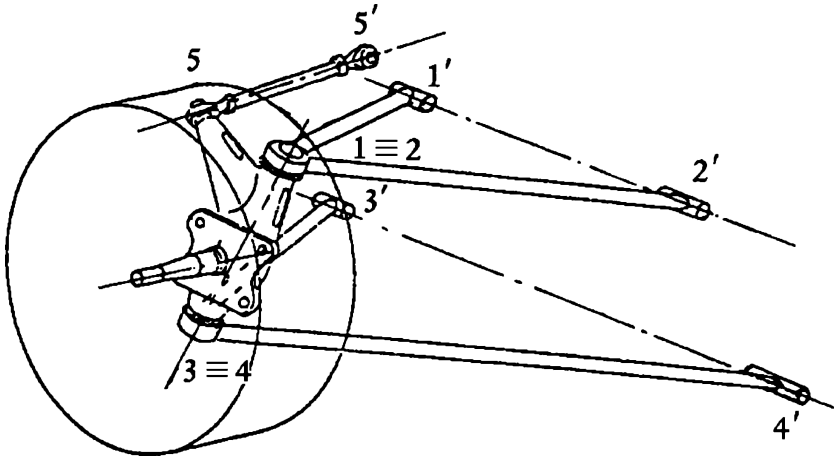


FIGURE 3.4. Low transversal quadrilateral suspension.

suspension, often referred to as an SLA (short-long arm) or A-arm suspension (Fig. 3.4). If lines $1'2'$ and $3'4'$ are parallel, the motion of the wheel is contained in a plane perpendicular to the line $1'2'$. The projection of the mechanism in such a plane is an articulated quadrilateral, whose side $1'3'$ is created by the vehicle body. A front suspension of this type is shown in Fig. 3.77, in which the engine is also sketched to show that this solution allows the mechanical parts of the vehicle to be located with a greater freedom than solutions based on rigid axles.

3.2.1 McPherson suspension

If the upper triangle is replaced by a prismatic guide a McPherson suspension is obtained (Fig. 3.5). Its simplicity and the fact that it leaves considerable free space for the engine has made it a common solution for automotive front axles, particularly in small cars.

The description of this suspension which is by large the most widespread configuration is divided into three parts.

The first part is addressed to defining suspension geometry and its kinematic behavior; a second will follow concerning the advantages and disadvantages implied by this suspension as compared to the remaining types.

This explanatory scheme will also be adopted for the other suspension types.

A third part is dedicated to the suspension and the design details of its most important components.

Similar components with easily imagined adaptations are also adopted in the remaining types. Therefore this section will not be repeated, unless for commenting on relevant differences with the McPherson suspension.

Description

The McPherson suspension is the most widely used one for front axles and is applied to all small and medium size cars. Some manufacturers also adopt this solution on large cars and sometimes sport cars as well; it is also sometimes applied to rear axles.

We are first going to describe wheel linkages. The wheel is guided, during its vertical motion, by means of a lower arm and a sliding guide, integral to the shock absorber. This device is also used as a spring installation. An upper pivot links it to the body.

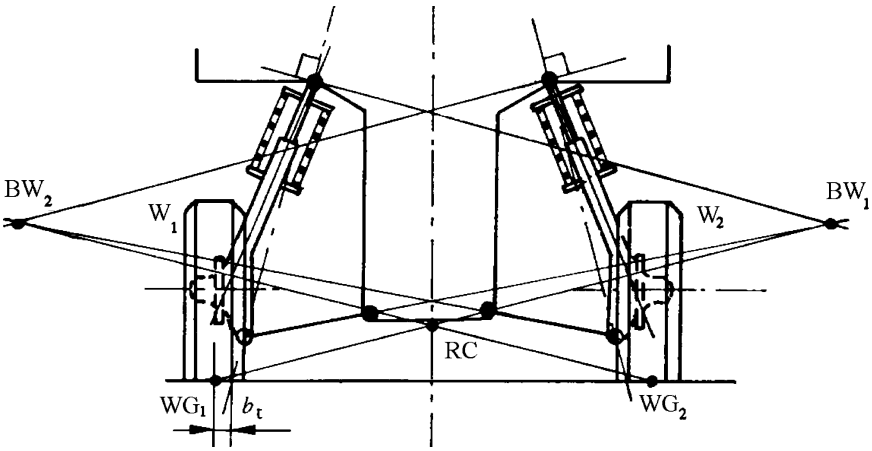


FIGURE 3.5. McPherson suspension. the figure shows the geometric construction to determine the roll center RC, not taking tire deformation into account.

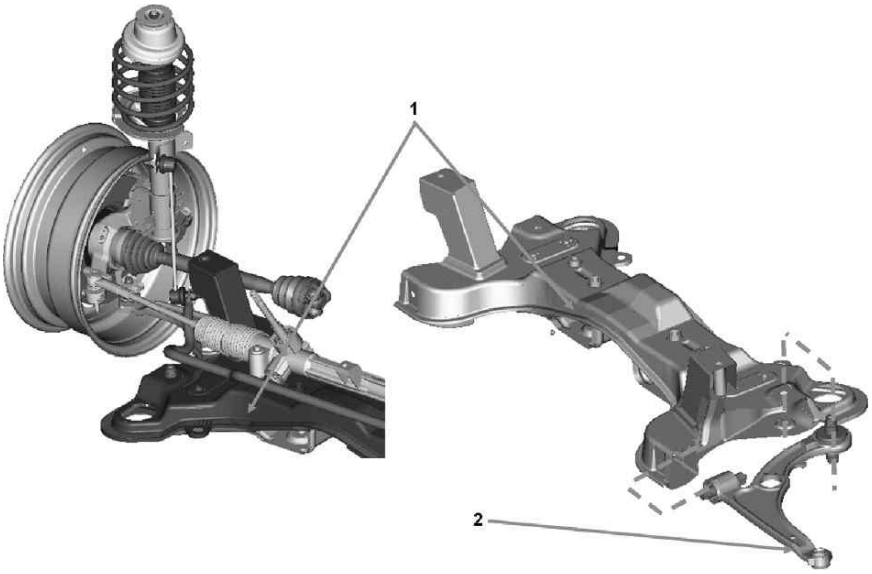


FIGURE 3.6. McPherson suspension for the front axle of a medium size front driven car; the lower arm (2) is articulated to the body through an auxiliary frame also called the subframe (FIAT).

We refer to a solution applied to the front axle of a medium size car, represented in Fig. 3.6, and proceed by describing the most relevant details.

The lower arm 2 is connected to the body through an auxiliary frame, also called *subframe* 1, at two different points, through elastic bushings 3. The lower arm is also linked to the strut 5 through a spherical joint 4, as shown in Fig. 3.7.

The strut is the element where the seat used to install the outer ring of the wheel roll bearing is machined. The inner ring is fit to a *hub*, which is flanged to the brake disc and wheel. On driving axles the hub is fixed to the drive shaft through a spline, allowing the transmission of torque from the differential to the wheel. On the strut two flanges are used to fix the brake caliper.

At the base of the body of the shock absorbers 6 (Fig. 3.8), two brackets 7 are welded, which are bolted to the strut in a rigid way. The spring rests on two seats, a lower seat 8 fixed to the shock absorber and an upper seat fixed to a thrust bearing 10; the upper ring of this bearing rests on an elastic mount, fixed to the body, in the wheel case.

The elastic mount is, therefore, the interface between the shock absorber piston and the spring, on one side, and the body, on the other side. The stop spring 13 is fit to the shock absorber piston and acts when the shock absorber tube touches its tip; this stop spring prevents direct metal to metal contacts in connection with large strokes and renders the elastic characteristic of springing more progressive.

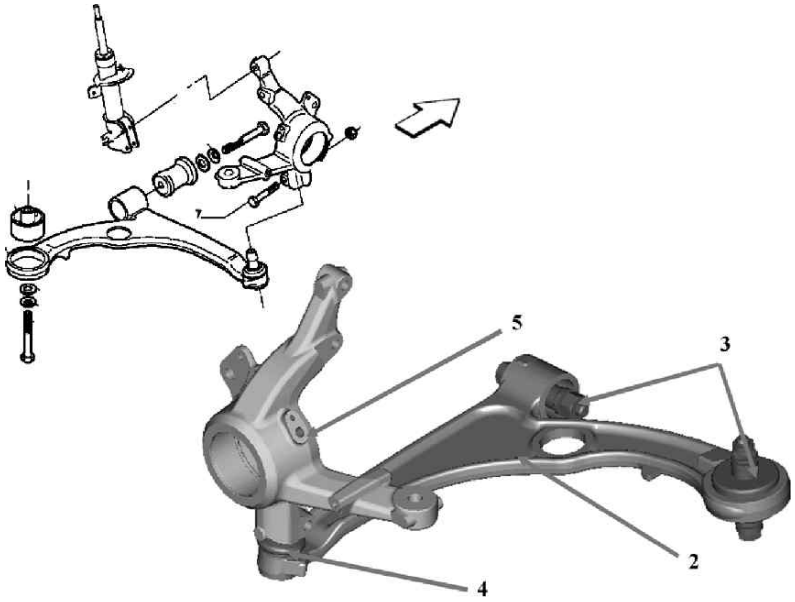


FIGURE 3.7. The lower arm 2 is connected to the subframe 1 in two different points, through elastic bushings 3; a spherical joint 4 links the arm to the strut 5.

The rack and pinion steering box 14 (Fig. 3.9), is bolted to the subframe and shows two steering tie rods 15, articulated to the rack through two spherical joints 16.

The anti-roll bar 17 is fixed on the subframe but is left free to rotate; it is connected to the shock absorber with two rods 18 (Fig. 3.10).

McPherson suspensions use the shock absorber piston to guide the wheel along the suspension stroke.

This detail ensures that external forces applied to the wheel at the contact point with the ground determine, in terms of the suspension geometry, a lateral force F_L and a momentum M applied to the piston itself.

Reactions R_A and R_B and the piston rod flexion have a major influence on shock absorber characteristics.

As will be explained, the shock absorber should exert a force ideally proportional to the suspension extension and compression speed; forces and flexion instead cause an almost constant friction that is independent of speed. This and other kinds of friction cause suspension *hysteresis*, which can be interpreted as the minimum force we must apply to the suspension to make it move.

Hysteresis is not desired because it causes the suspension to block when forces applied are lower than a minimum value. This characteristic of the McPherson suspension must be acknowledged as an inconvenience.[3]

[3]In reality, this inconvenience is not only present in McPherson suspensions but on all suspensions where shock absorbers are partially loaded by wheel forces.

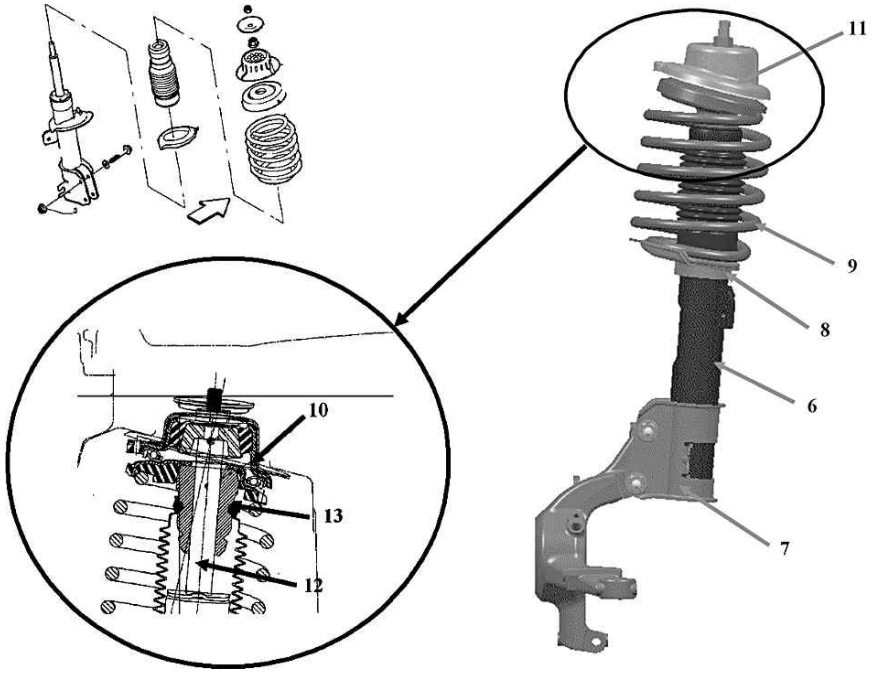


FIGURE 3.8. At the base of the shock absorber 6 two welded brackets 7 are used to fix the strut in a rigid way; the spring 8 is contained between two seats, one connected to the shock absorber, the other to the thrust bearing 10; the upper ring of this bearing rests on the elastic element 11, fixed to the body.

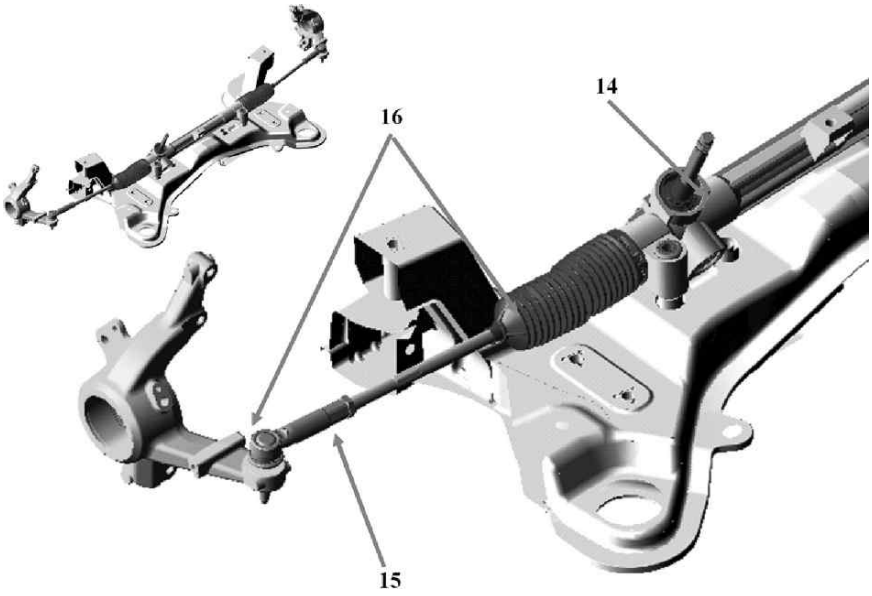


FIGURE 3.9. Rack and pinion steering box mounted on the subframe of a front McPherson suspension.

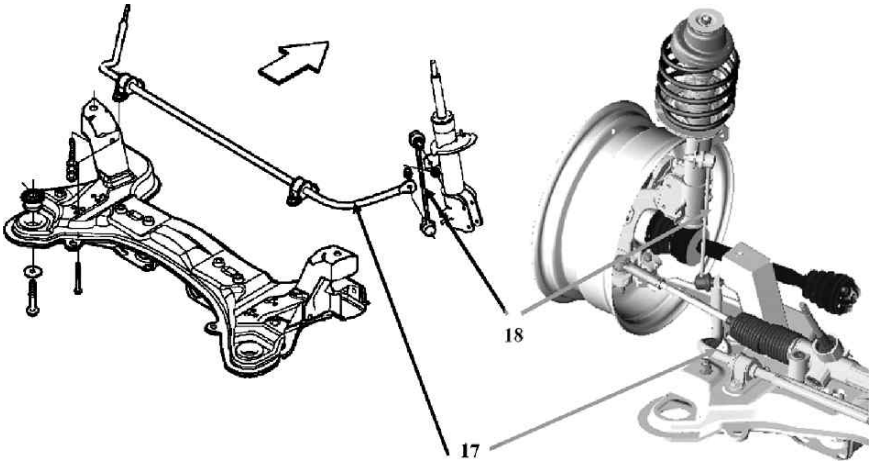


FIGURE 3.10. The anti roll bar 17 is fixed on the subframe, but is left free to rotate; it is connected to the shock absorber with two rods 18.

With reference to Fig. 3.11, reactions R_A, R_B and deflections of the piston rod of the shock absorber ϑ_A and ϑ_B can be calculated using the following formulae.

It should be noted that the fictitious reaction R_T is the sum of the absolute values of reactions between piston rod and its bushing (R_A) and between piston and cylinder (R_B); this force multiplied by the friction coefficient (considered to be the same in the two couples) will show, neglecting the contribution of rod deformations, the friction force that reacts to rod motion, that is the hysteresis.

We define:

E the elastic modulus of the piston rod;

I the flexural moment of inertia of the rod cross section.

We can write:

$$\begin{aligned} R_A &= \tfrac{1}{b}\left[F_L\,(a+b) - M\right] \ , \\ R_B &= \tfrac{1}{b}\left(F_L a - M\right) \ , \\ R_T &= |R_A| + |R_B| \ , \end{aligned} \tag{3.1}$$

$$\begin{aligned} \vartheta_A &= \tfrac{b}{3EI}\left(F_L a - M\right) \ , \\ \vartheta_B &= \tfrac{b}{6EI}\left(F_L a - M\right) \ . \end{aligned} \tag{3.2}$$

Let us consider now the possibility of introducing an offset e between the rod and the spring symmetry axis. This offset will imply the application of a moment M, exerted on the rod by the spring, as shown on the figure, that will reduce reactions acting on the bushings. If:

$$M = F_L a \ , \tag{3.3}$$

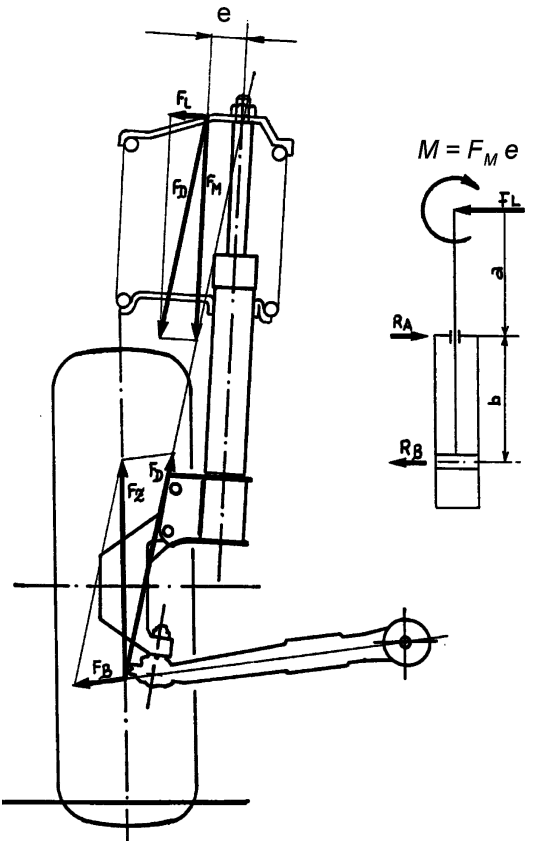


FIGURE 3.11. Geometric scheme useful for calculating forces on the sliding bushings of a shock absorber of a McPherson suspension.

we obtain the minimum value for R_T under these conditions:

$$\begin{aligned} R_A &= R_T = F_L \ , \\ R_B &= 0 \ , \end{aligned} \tag{3.4}$$

$$\vartheta_A = \vartheta_B = 0 \ . \tag{3.5}$$

Magnitudes M and F_L change according to thrust bearing architecture and spring inclination axis.

With reference to Fig. 3.12, we identify two different families of McPherson suspensions: Those with integral thrust bearing (left) and those with double thrust bearing (right).

In the integral solution, where both spring and shock absorber loads are exerted on the body through a single rubber element, reactions on the rod are represented by the formulae we have examined.

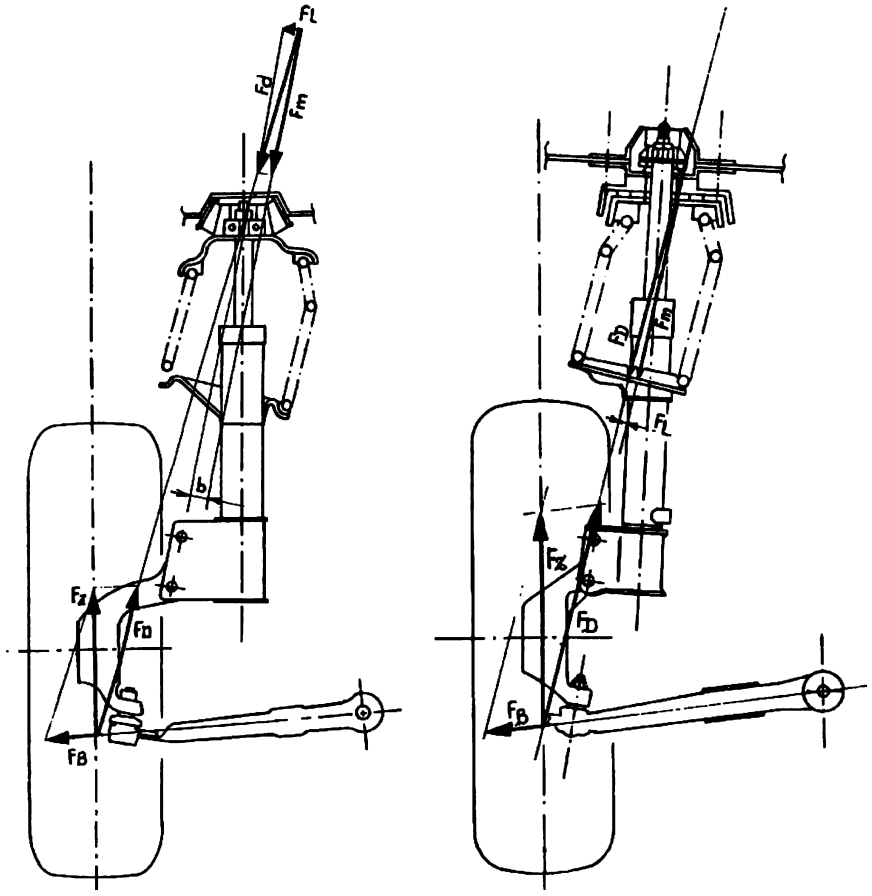


FIGURE 3.12. Scheme of McPherson suspension with integral (left) and double (right) thrust bearing.

In the double solution, spring load and shock absorber load are transferred to the body through two different rubber elements and the force acting on the rod is represented by F_L only.

In both cases is possible to reduce the hysteresis force, applying a suitable offset to the spring.

The zero condition for hysteresis is bound to the value we have chosen for the vertical load; changing this condition or applying other loads, as for example a cornering force, will cause a zero condition to change. Spring offset can be designed for a single load condition only, judged to be statistically significant.

Advantages and disadvantages

This comparison of advantages and disadvantages refers to front axle suspensions.

Advantages

- Design simplicity and reduced cost.
- Because of the relevant separation of body joints, forces exerted on the body are low in comparison, for example, to a low double wishbone suspension.
- Higher suspension stroke than in other suspensions (a high double wishbone one for example, because of the limitation on upper arm length).
- Contained transversal dimension, due to the absence of the upper arm; this fact is quite beneficial for transversal engine installation.
- Possibility of designing with superior longitudinal flexibility, without greatly affecting the caster angle.
- Freedom in designing elasto-kinematic properties; camber recovery is limited only by viable positions for the upper pivot and lower arm fixed joint.
- The ratio between suspension and shock absorber stroke is near to one. Shock absorbers therefore work well with limited loads, low oil heating and valve wear.

Disadvantages

- Lower performance in camber recovery. See, for example, the comparison between camber angle variation for a McPherson and a double wishbone suspension, shown in Fig. 3.13.
- Suspension characteristic geometry causes a position for the upper pivot interface with the body, usually called *dome*, which is usually far removed from the stiffest structures of the body, the side beams, as shown in Fig. 3.14. This causes significant problems with suppression of vibrations and noise from the road.
- Shock absorber piston rod deformation can increase friction and hysteresis.
- Notable height for the upper pivot, so that the spring and shock absorber are set over the wheel; this fact could degrade the vehicle's aerodynamic shape and sporty body style.

Design details

Let us examine now some of the most relevant components of McPherson suspensions, which can feature many alternative versions.

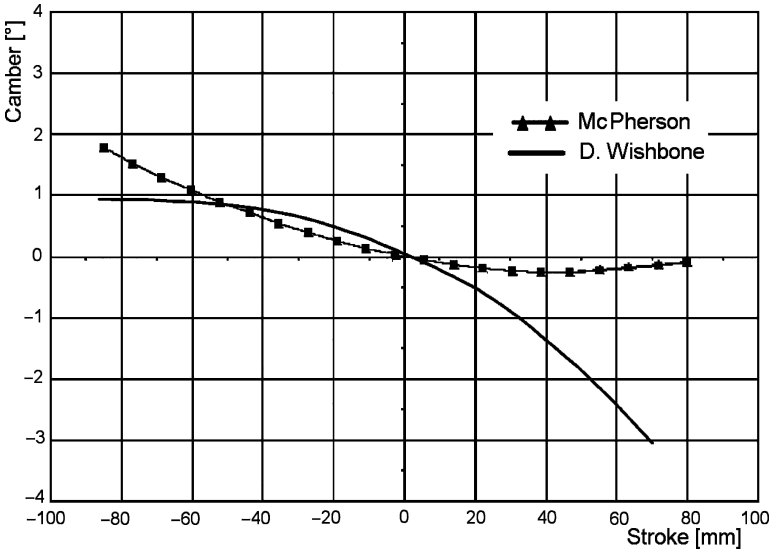


FIGURE 3.13. Comparison between camber angle variation as a function of suspension stroke in a McPherson and a double wishbone suspension; the difference produces a better camber recovery in the double wishbone suspension.

Lower arm

Lower swinging arms connect moving parts such as strut and wheel directly to the body or the subframe.

The lower arms have a particular shape allowing both suspension and steering motions to the wheel, as shown in Fig. 3.15, and are connected to the subframe through elastic bushings and with spherical joints to the strut.

The choice of the shape of the lower arm is part of the concept design of the front suspension and must be compatible with handling and comfort targets.

Once the suspension scheme has been decided, using a CAE tool, the envelope of the external shape of the tire can be designed, factoring in steering motion and suspension stroke, as shown on the right side of Fig. 3.15. The lower arm shape must obviously be outside this envelope and must take into account assembly and disassembly conditions (holes must sometimes be provided for access by screwdrivers and wrenches) and limitations imposed by other parts of the car, such as the power train or body.

The lower arm absorbs a relevant portion of the forces coming from the wheel during braking or cornering maneuvers. Because this component is fitted to the body through articulations, reaction forces must be contained in the plane through their centers.

The lower arm oscillates around the axis through the two bushing centers (points E and F of Fig. 3.15; remember the position of this axis with reference to

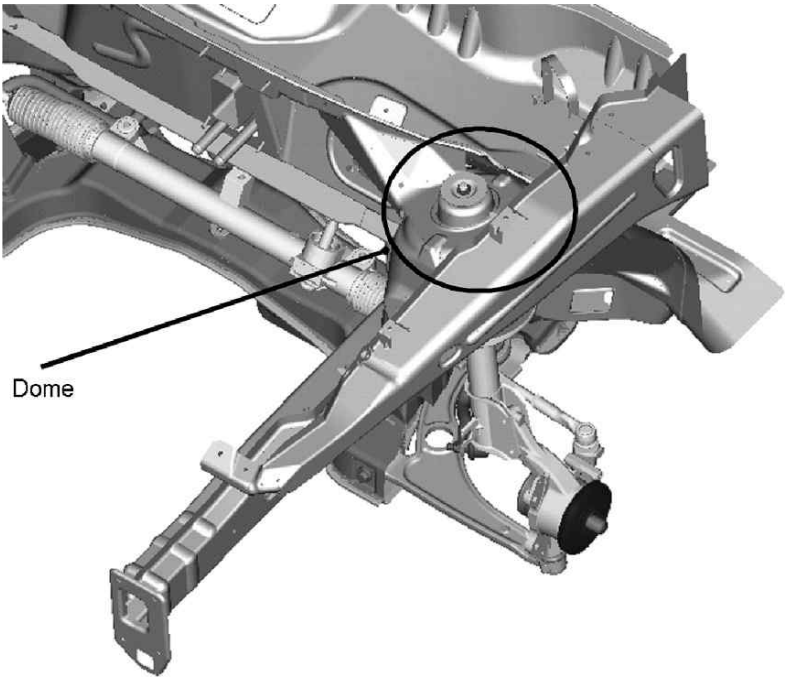


FIGURE 3.14. The position of the dome imposed by McPherson suspensions is usually far removed from the stiffest body structures, which are lower, close to the side beams. (FIAT).

the body). From a kinematic point of view the A point trajectory is not affected by the position of points E and F along their axis.

The axial position of the bushing is, instead, relevant for the magnitude of the reaction forces at points E and F. Bushings are not rigid spherical joints. They are made with elastomeric material and are deformed according to the amount of force they receive.

The shape of the bushing governs its flexibility: As we have already said, suspension displacement under the applied loads is relevant for handling and the comfort characteristics of the vehicle.

These kind of considerations have led car manufacturers to design lower arms in a banana-like shape (Fig. 3.16).

Because of this shape (Fig. 3.17, at left) the most important part of the lateral force applied to the A point is loading bushing E. This last must be designed very stiff so as to limit camber variation (opposite to recovery of camber) under the effect of such forces.

For the same reason the moment of longitudinal forces around point E loads the bushing F transversely.

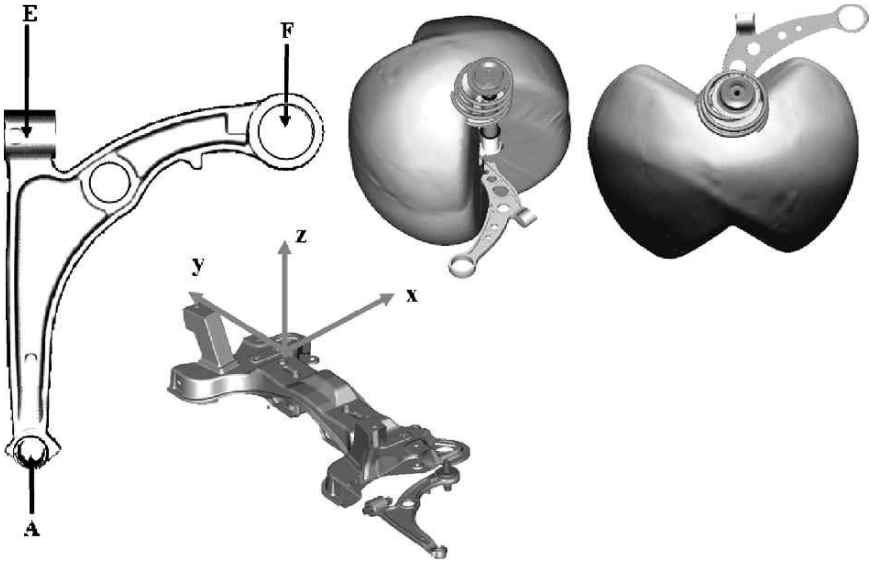


FIGURE 3.15. The lower arm must allow both stroke and steering motions of the wheel, as shown by the envelope shape on the right; articulations are made by spherical joints or elastic bushings according to their function.

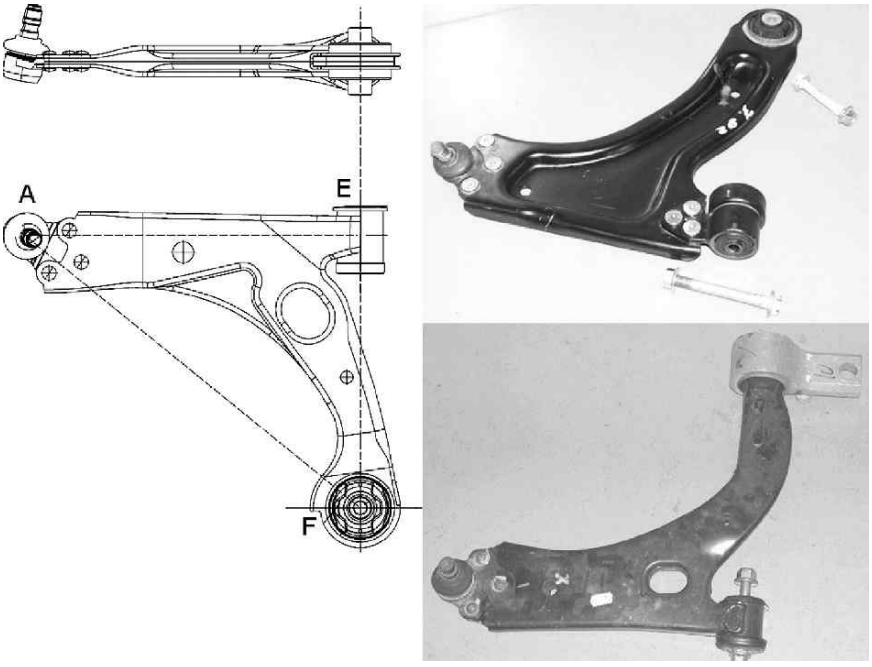


FIGURE 3.16. The banana like shape of the lower arm of the McPherson suspension is the result of a compromise between handling and comfort.

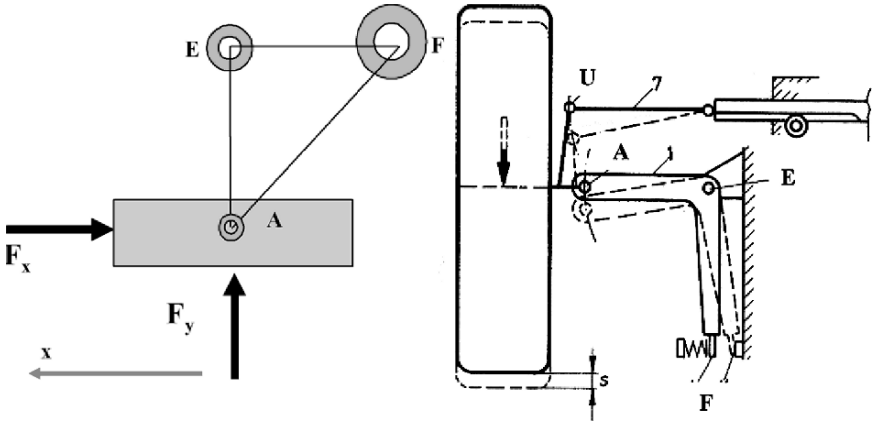


FIGURE 3.17. The greater part of lateral force through A is applied to point E; this bushing is designed to be stiff and to limit camber variation, because of the external load. Bushing F is designed to be flexible to allow the arm to have a compliant displacement when the wheels encounter obstacles.

Bushing F must show a high level of flexibility. The lower arm can then rotate around the rigid bushing E. As a consequence, when the wheel encounters an obstacle it can have a compliant back displacement, with a beneficial effect on comfort.

But point A also determines the king-pin axis position, together with the upper pivot. To limit the toe angle variation derived from longitudinal motion of the wheel, steering arm 7 should be suitably positioned, so as to shape a parallelogram like that seen in Fig. 3.17, at right.

On some cars the lower arms are shaped like a isosceles triangle, where point A is the vertex; in this solution bushings E and F are almost equal in terms of dimensions and stiffness. In this case the target is to obtain superior longitudinal flexibility without affecting toe angle variations negatively. This scheme must be adopted with a different shape of subframe, one placed in a more advanced position than that used for banana lower arms.

Lower arms can be made by cast iron, by hot stamped steel or by cold stamped steel sheets; if weight reduction is a priority stamped aluminum can also be applied.

In Fig. 3.18 two examples are shown: One in cast iron and one in stamped steel. The iron technology achieves minimum cost, but these components feature reduced mechanical characteristics in terms of rupture percentage elongation.

Lower arms made by stamped steel sheet are built welding two semi-shells together. In this case the spherical joint housing has to be riveted, while bushing housings show local reinforcements. This solution is characterized by reduced weight, better rupture elongation, and slightly increased cost.

Figure 3.15 shows an example of hot stamped aluminum arms; the shape of a hot stamp steel arm would be similar except for the notable reduction in cross

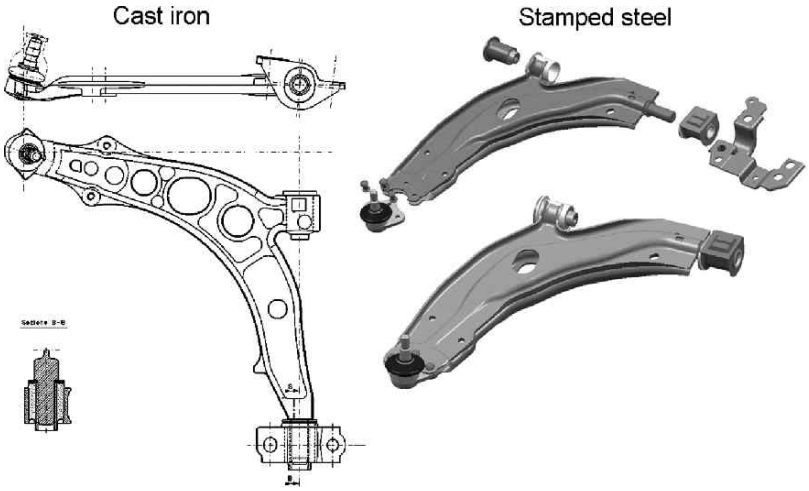


FIGURE 3.18. Examples of lower arms made by cast iron or stamped steel sheet, in two shells to be welded together.

section. With this technology excellent mechanical properties can be obtained at a higher cost; they allow notable weight reduction, particularly the aluminum solution.

All these solutions must be machined to prepare the mounting interfaces for the spherical joint and bushings with the necessary dimension tolerance.

The importance attributed to the percentage rupture elongation comes from consideration of accident shocks that the suspension might experience. Let us consider, for example, the case of an accidental side shock of the wheel against a side walk, caused by excessive speed in a curve.

In this case the lower arm has the function of a sacrificial element, defined by its rupture properties, which include load, deformation and absorbed energy.

It should be noticed that in some cases the anti-roll bar is directly linked to the lower arm; a suitable interface on the arm must then be provided.

During an asymmetric suspension stroke, the lower arm is also loaded by bending moments, coming from the anti-roll bar, that also involve the bushing and spherical joint.

Elastic bushings

Elastic bushings with their properties of deformation allow motion of the arm with reference to the car body or subframe.

They are made by an outside body 3 and inside body 1, respectively fixed to the suspension arm and the body or subframe; between these two bodies a layer of rubber 2 is vulcanized.

Figure 3.19 shows two different installation alternatives. On the left side solution A is shown, where the outside body is flanged on the subframe and the

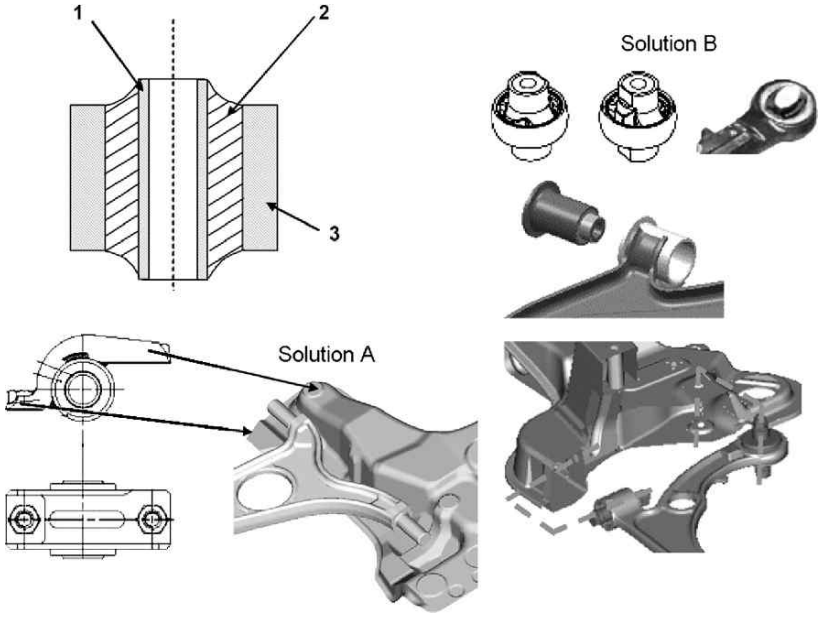


FIGURE 3.19. Elastic bushings are made by an outside body 3 and inside body 1, respectively fixed to the suspension arm and the body or subframe; between these two bodies a layer of rubber 2 is vulcanized. Note the two different installation solutions.

inside body is press fitted on the arm. On the right side solution B features the outside body press fitted into the arm and the inside body flanged to a fork in the subframe.

Rubber thickness and mix composition are designed to be able to vary stiffness according to the load direction. Vulcanized rubbers used in these applications must have a high fatigue limit.

We have seen in the previous paragraph that bushings must be designed to grant the suspension a certain elasto-kinematic behavior in terms of toe angle and wheelbase (longitudinal displacement) variations.

Bushings with high stiffness will result in superior road holding or handling behavior. In this case the layer of rubber between the two steel bodies will be a few millimeters thick and the bushing outside diameter will be in the range between $35 \div 45$ mm.

The need for high flexibility in bushings producing superior comfort requires thicker rubber; therefore the outside diameter of the bushing could reach the value of about 60 mm.

Figure 3.20 shows, as an example, the elastic characteristics of the more flexible bushing of a McPherson suspension. The stiffness in the x direction and in the y direction are very different; for this reason the rubber body is not made like a solid ring, but presents a connection, between the inside and outside bodies,

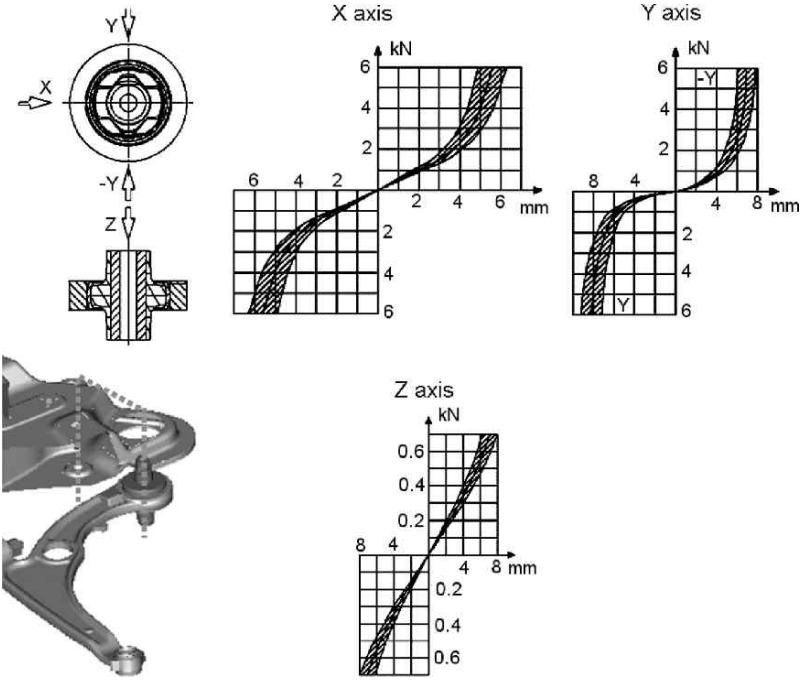


FIGURE 3.20. The stiffness of the more flexible bushing is very different in the x and y directions; for this purpose the rubber interior is designed as a beam, being bent in one direction and compressed in the other, in order to produce an asymmetric stiffness.

made like a beam, bent in the y direction and compressed in the x direction. Two stop springs made of rubber limit the bending deformation.

In a bushing the following values of stiffness have to be defined:

- Radial stiffness along the two x and y directions of the body reference system, when the bushing internal diameter axis is, as in this case, directed to the z axis
- Axial stiffness, along the z axis
- Conical stiffness, measured by imposing on the bushing axis assigned angular deviations, with reference to the zero load condition
- Torsional stiffness, measured by imposing relative angular rotation of the steel body around their axis

These values of radial stiffness must be recorded in drawings, in addition to geometric and material information, to allow production quality control; for this purpose it is necessary to show these elastic characteristics with an allowed tolerance band.

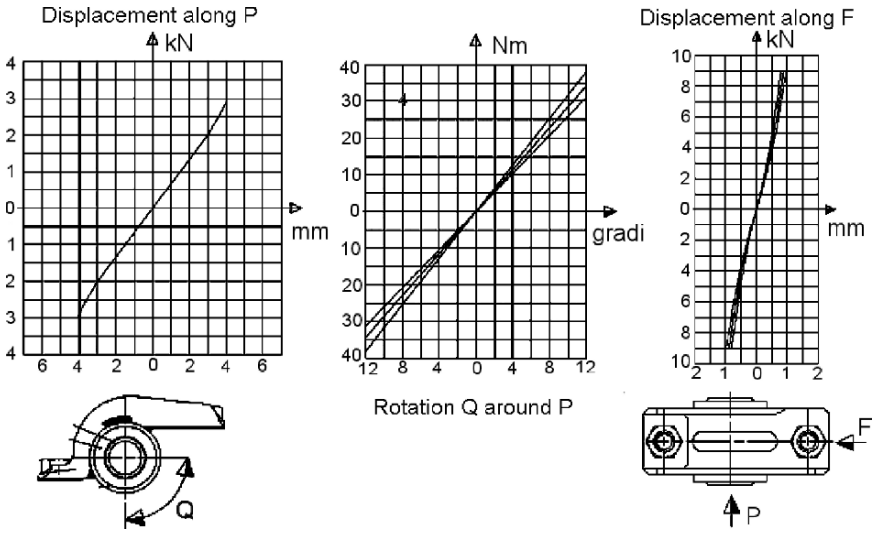


FIGURE 3.21. These diagrams show the elastic characteristics of a stiff bushing, engineered for road holding behavior.

Conical and torsional stiffness control is not as critical, because their impact on elasto-kinematics is limited; these values must be small, in order to allow a free arm motion. It is, instead, important to verify that the arm motion, caused by the suspension stroke, does not affect fatigue life or change the elastic behavior of the bushings at high deformation angles.

Figure 3.21 shows, instead, the elastic characteristics of a stiff bushing, engineered for road holding and handling performance.

Many cars also feature hydraulic type bushings. These can add to the described elastic characteristics a damping characteristic, produced by squeezing a quantity of oil through internal orifices.

These damping properties can have a beneficial effect in smoothing a vehicle's ride.

Spherical joints

The articulation of suspension strut with lower arm is made by a spherical joint also called a *steering knuckle*. Other spherical joints are used at the end of the steering bars, connecting steering rods with steering rack.

This kind of joint is stiff in every direction but the radial and allows high rotation angles of the connected elements.

Figure 3.22 shows the spherical joint with its spherical head and a fixing pin. In solution A a screw is used to press a hole cut in the strut around a cylindrical pin; in solution B the pin is tapered and pressed into the strut hole by a traction nut.

The choice between the two alternative solutions has no precise rules, but falls back on tradition or existing tools for machining and assembling.

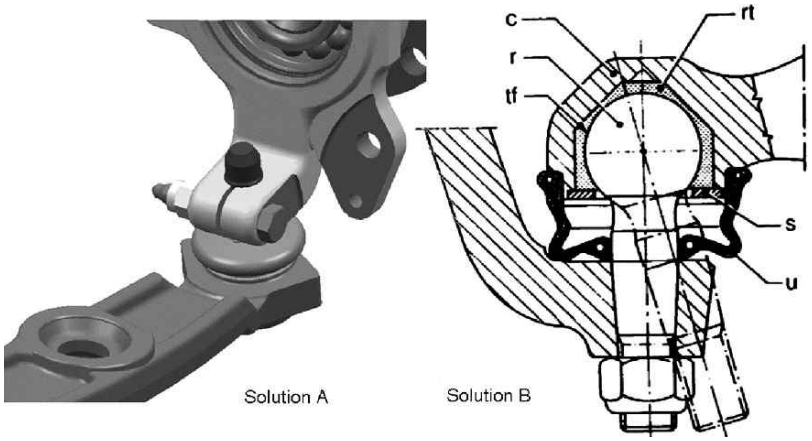


FIGURE 3.22. The spherical head of the joint has a fixing pin. In solution A a screw is used to press a hole cut in the strut around a cylindrical pin; in solution B the pin is tapered and pressed into the strut hole by a traction nut.

In both cases the material of the sliding surface of the spherical cavity is fixed into the arm directly; it is fixed inside the arm by plastic deformation. A layer of pre-lubricated Teflon is set between the spherical head and its mounting cavity to lower the friction during rotation.

It is a good practice to prescribe in the drawings, in addition to the geometric dimensions and materials, the maximum working angles to be allowed and the minimum radial stiffness assumed to be necessary.

Strut

The strut is connected with the majority of suspension components and transfers to them any force coming from the wheel. Its tasks condition its geometry. The severity of the applied loads characterize the strut by necessitating its squat shape with thick cross sections.

We can identify the following functions of the strut (with reference to Fig. 3.9):

- To hold the wheel bearing in both radial and axial directions
- To connect with the suspension arm through the spherical joint
- To offer the fixing flange for the brake calipers
- To fix the brake cover
- To fix the lower end of the spring and shock absorber unit
- To connect with the steering arm through a spherical joint

If the car is supplied with an ABS system or a vehicle dynamics control system, the strut is also use for fixing the wheel speed sensor.

Materials used for this component include hot stamped steel and cast iron; sometimes hot stamped aluminum is used in luxury cars.

Hub and bearing

To allow hub rotation and to transfer the wheel forces to the strut, the hub is mounted with a double range ball bearing or a double tapered roller bearing; tapered roller bearings are applied with opposed conicity.

Figure 3.23 shows the hub assembly for a front wheel driven vehicle with ball or tapered roller bearings. This bearing configuration, specific to cars, is characterized by bearings with a single outside ring and double inside rings.

This is commonly called the O configuration; the pre-load of the two bearings is determined by the play allowed between the two inside rings, which is brought to zero after mounting.

The solution used on most small and medium size cars provides a double range ball bearing, as we have seen until now. The hub pre-load is obtained by tightening the nut of the hub at a certain torque. This kind of bearing is also called first generation.

Second generation bearings (Fig. 3.24) still provide two inside rings, but the outside ring is directly integrated into a flange to be screwed to the strut. This solution simplifies the strut machining operation to cut the seat for the bearing. It also reduces the assembly cost; on the other hand these bearings are themselves more costly.

Third generation bearings feature a single inside ring, including the wheel hub, in addition to the mounting flange on the outside ring; this solution has

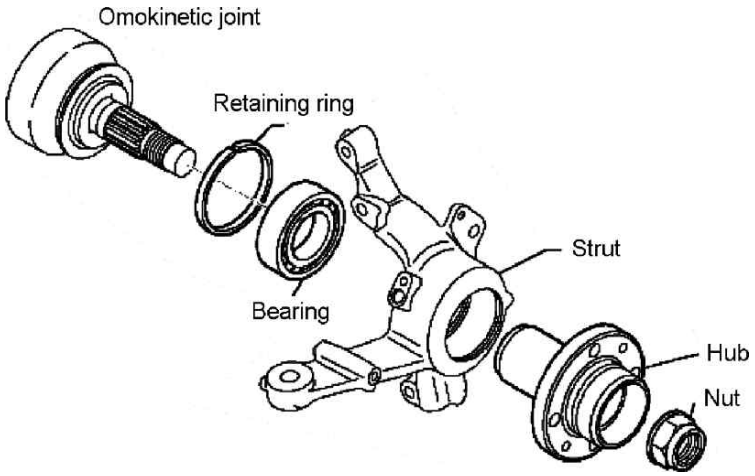


FIGURE 3.23. Hub assembly of the front wheel for a front wheel driven vehicle.

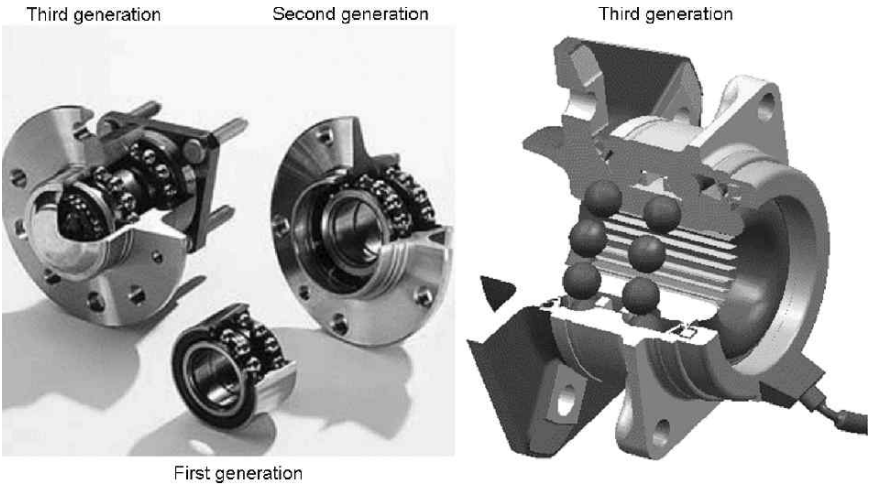


FIGURE 3.24. Wheel bearings of first, second and third generation.

the advantage of maximum reduction in space and weight and of having a sealed bearing; the cost of the bearing is, however, much increased.

Coil spring and stop spring

A coil spring alone would give the wheel an almost linear elastic characteristic, considering that the relationship between wheel and spring stroke is almost linear.

As will be explained in the second volume, in the chapter on comfort, a linear characteristic should be avoided because spring deformation should decrease in proportion to the load increase. This target is reached by applying additional springs made of rubber. These are called *stop springs* because their initial purpose was to limit the suspension stroke to a certain value.

The progressive spring characteristic, added by the stop spring, can keep the first natural frequency of the sprung mass at a constant value, even if the vehicle payload is changing, while limiting the compression stroke as well; another stop spring limits the extension stroke and affects the spring ratio.

When the shock absorber is used for the coil spring installation, as in McPherson suspensions, the stop spring too must be fitted into the shock absorber assembly. Figure 3.25 shows the example of a rear McPherson suspension.

In front suspensions differences will be minimal except for the lower mount. The compression stroke stop spring is coaxial with the shock absorber piston rod in the outside of the hydraulic cylinder; the extension stroke stop spring is much smaller and is likewise coaxial with the piston rod, but inside the hydraulic cylinder.

In the same figure a detailed drawing of the stop spring is shown; these springs are usually made of polyurethane. In the middle of the same figure the

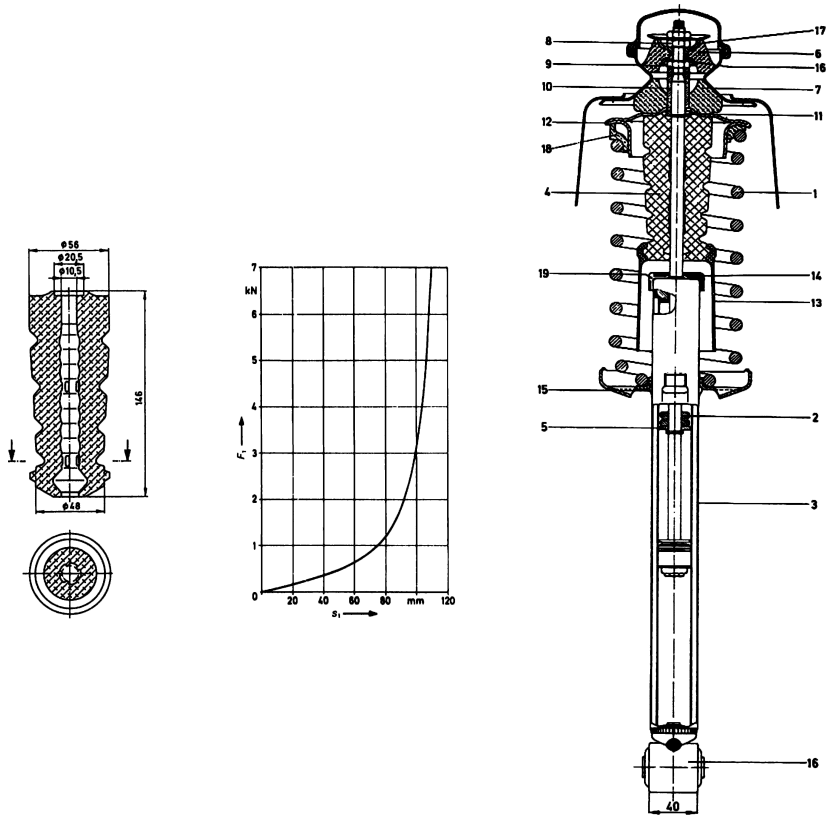


FIGURE 3.25. The stop spring 4 involving the compression stroke is coaxial with the shock absorber piston rod, outside of the hydraulic cylinder; the extension stroke stop spring 2 is much smaller and is found inside the hydraulic cylinder.

elastic characteristic is shown; the shape is strongly nonlinear and is obtained by shaping the tip of the spring like a cone. Ring grooves also contribute to this characteristic.

In suspensions where coil springs and shock absorbers are not unitized, the stop springs can be separated; design criteria remain the same.

The extension stroke stop spring is almost always inside the shock absorber.

The main spring is usually of the coil type; spring retainers are made by stamped steel sheet and are shaped so as to offer a wide supporting surface to the coil; free coils can have constant or variable pitch. Coils coming into contact progressively increase the total coil stiffness. The coil spring alone can feature the desired progressive elastic characteristic.

Anti-roll bar

The roll angle resulting in a curve, because of the suspension elastic characteristic given by normal springs, could be excessive. In this case an additional anti-roll bar is used.

This is an additional spring that increases suspension rigidity only by antisymmetric strokes, as in a curve, where one suspension is compressed and the other is extended. If both suspensions are moved for the same stroke, in the same direction, this additional spring has no effect.

In our example, the anti-roll bar (Fig. 3.26) is made up of a bent tube connected to the subframe in two points. The tube ends are flattened and drilled, to be connected to a pendulum bar that is linked to the shock absorber at the other end.

During body roll the wheel to the outside the curve is compressed, while the inside wheel becomes extended. The body motion torques the linear part of the bar which tends to limit the difference in strokes, limiting in this way the angle of roll.

The pendulum bar (see Fig. 3.27) limits force exchange with the suspension to the anti-roll torque and avoids bar stressing during symmetric strokes; this happens if the bar is flanged directly to the suspension arms, as in some small cars.

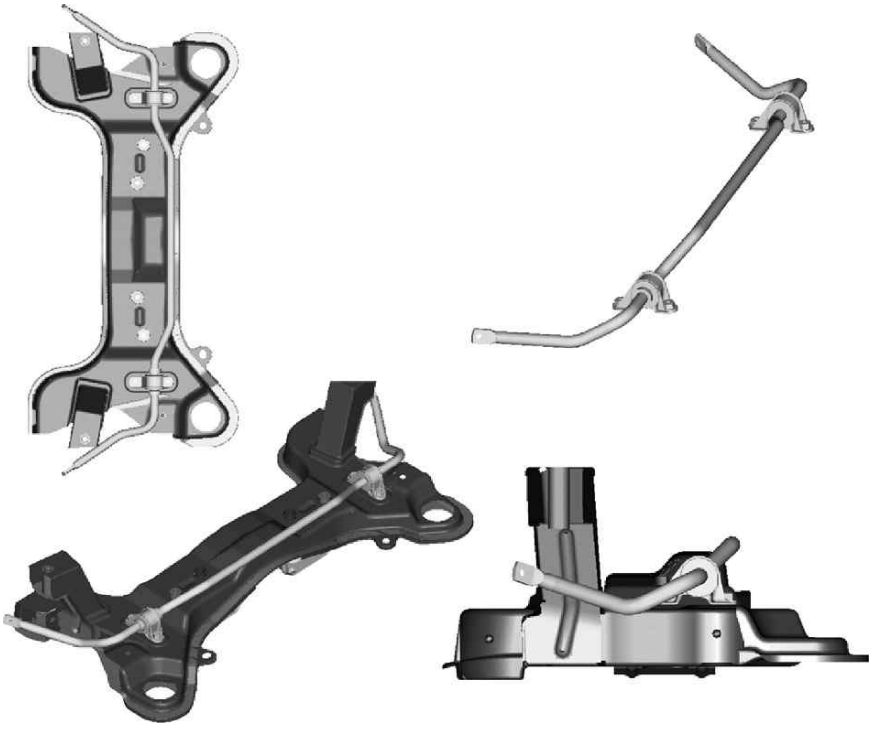


FIGURE 3.26. Anti-roll bar mounted on the suspension subframe.

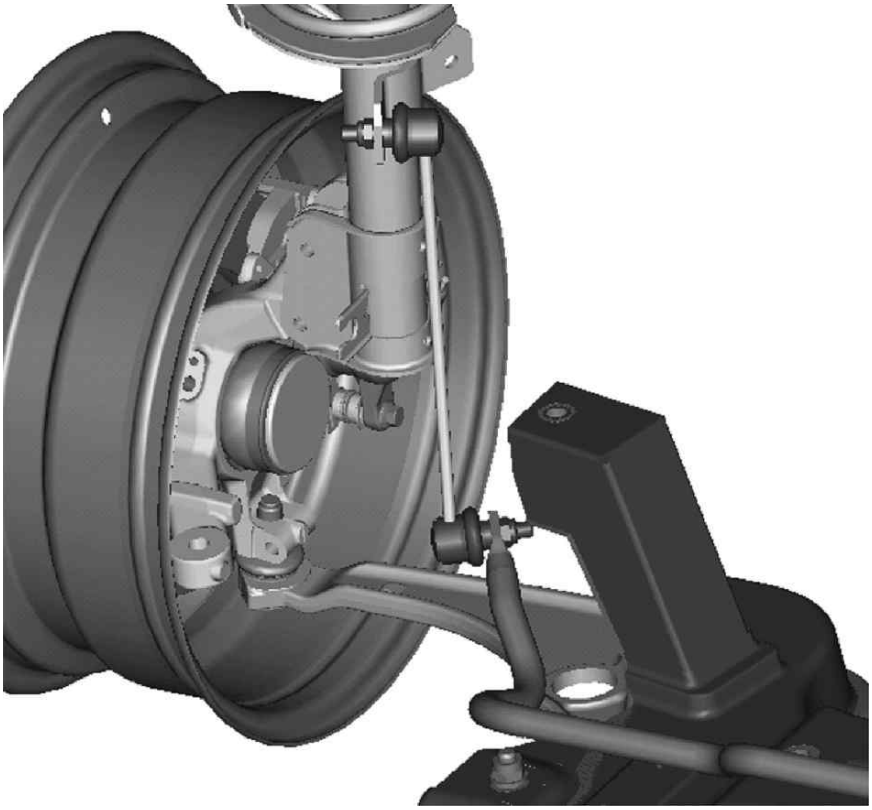


FIGURE 3.27. Anti-roll bar connected to the shock absorber through a pendulum bar.

In this last case, in fact, as shown in Fig. 3.28, the interface between bar and arm is required to follow the circular trajectory of the arm. This fact introduces an additional bending stress to the anti-roll bar; this arrangement has the further disadvantage of an unfavorable ratio between the bar torsion angle and the suspension stroke, as compared with the pendulum bar application. The result is a reduction in the bar effect.

The pendulum bar can also be connected to the suspension arm, usually at a shorter length; in this case arm swinging should not introduce additional deformation to the anti-roll bar but only the pure torsion, so as not to affect comfort by symmetric strokes.

Figure 3.28 also shows on section A-A the elastic bushing mounts of the anti-roll bar, attaching it to the sub frame; these mounts are common to any kind of bar.

Shock absorber

The shock absorbers used on vehicles today are solely hydraulic; the damping force is obtained by squeezing with a piston a certain amount of oil through

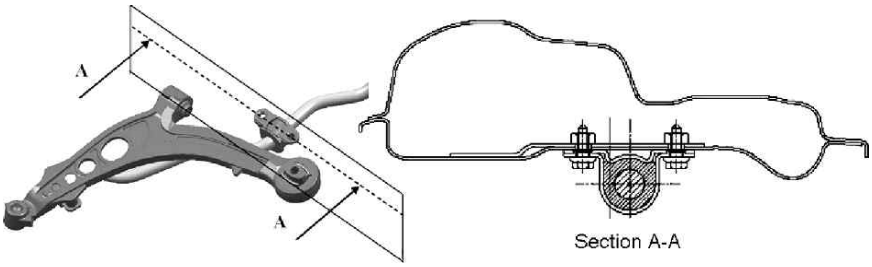


FIGURE 3.28. Anti-roll bar directly fixed to the suspension arm.

small orifices controlled by valves. The flow of oil through the orifice occurs, at a given area of the flow cross section, with a certain pressure drop that is almost proportional to the flow rate.

Shock absorbers can be classified as of the *structural* or *conventional* non-structural type, according to their capacity for supporting loads perpendicular to the piston rod.

The first type is used on McPherson suspensions.

The shock absorber is fit to the suspension strut and carries part of the load coming from the contact point of the tire; these loads generate on the piston rod shear forces and bending moments as we have seen in Fig. 3.11. These have consequences for piston bar diameter and sliding bushings.

The second type can be found on almost all remaining types of suspension, where the shock absorber receives forces only along its piston rod.

Finally, shock absorbers can be unitized with the coil spring or separated; there are McPherson suspension where the two elements are separated, as there are other type of suspensions where they are unitized, with benefits for installation space and ease of assembly.

In conventional shock absorbers the only load comes from the internal pressure that loads, by definition, the piston rod.

It should be noticed, however, that transversal forces could also be applied, causing an increase in hysteresis and additional stress on the piston bar; this situation may occur when, because of the stroke of the suspension, the shock absorber changes its inclination. In this case, the torsional or conical stiffness of the elastic bushings connecting the shock absorber to the suspension and the car body can introduce transversal forces; this can also happen with sliding bushings, because of their internal friction.

Figure 3.29 shows on the left a scheme useful for understanding the operation of a hydraulic shock absorber. The upper part of the shock absorber is fixed to the sprung mass, the lower to the unsprung mass; if the shock absorber has no structural function the joints should be equivalent to a spherical joint.

During its stroke the suspension moves the piston rod 1 and the piston 2, inside the cylinder 3. The volumes outlined with broken lines A, B and C are filled with oil; the piston motion, because of the incompressibility of the oil, squeezes a certain flow rate from chamber A, to chamber B, through valves V_1 and V_2.

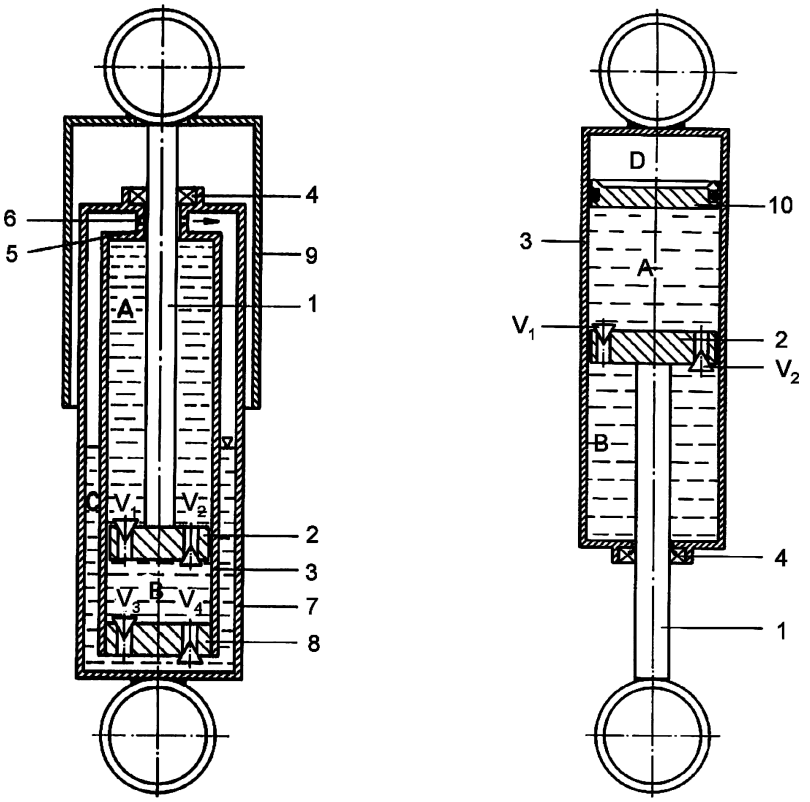


FIGURE 3.29. Hydraulic scheme for understanding the operation of a double tube (left) and mono tube (right) shock absorber.

From this flow a damping force is generated; it should be noted that valves V_1 and V_2 are outlined as check valves. It is possible in this way to obtain different damping coefficients during the compression stroke (valve V_1) and extension stroke (valve V_2).

The oil pressure generated by friction losses through the valves must be tightened by the seal 4 to avoid oil spillage; this seal causes a certain friction and suspension hysteresis.

To improve seal effectiveness an additional labyrinth 5 is provided, reducing the pressure acting on the seal to a value close to the ambient pressure of a chamber connected to the holes 6.

The penetration and the extraction of the piston rod must be accompanied by a displacement of an oil volume equal to that occupied or freed by the rod.

In our scheme, volume C, with free space contained in the cylinder 7 provides this function through the check valves V_3 and V_4, communicating with the volumes A and B. The protection tube 9 avoids water or dust contamination on the rod; this protection can be improved with a rubber bellows, not shown in this figure.

This type of shock absorber is also called *double tube* because of the double outside cylinder. Two chambers are therefore available (inner and outer) containing oil. The piston with the valves assembly enters the inner chamber, while a second valves assembly separates the inner from the outer chamber.

In the lower part of the outer chamber, a quantity of pressurized gas compensates for the oil volume variations created by the piston motion. The inside pressure of the compressed gas avoids cavitation near the valves, where pressure drops are present.

Valve assemblies have different functions, according to the direction of flow. During suspension extension the piston valves control the damping force, while the lower valves allow the compensatory flow between the inner and outer chambers. During suspension compression the lower valves have the damping function, while those on the piston allow an almost free flow.

Some of the orifices are closed by pre-loaded valves that open when the pressure drop exceeds a certain limit. In this way the damping force can be limited for high compression or extension speeds.

A second kind of shock absorber is called *mono tube*, with reference to the scheme on the right of the same figure. This kind of shock absorber is longer, but with a smaller diameter as compared with the double tube type.

Only two oil chambers are available, A and B in the scheme. The valve assembly on piston 2, with rod 1, controls the damping force in both directions.

In particular, the check valve V_1 controls the compression force, while V_2 controls the extension force. A volume of pressurized gas D is separated from the oil through a floating piston 10. This gas volume compensates for oil volume variations due to piston rod motion.

The advantage is a simplification of the design, while the disadvantage is an increase of pressure on the seal 4, with potential leakages and increased hysteresis. The pressure in chamber D applies to the suspension a non-negligible force, in parallel to the coil springs.

The force-speed characteristic of a vehicle shock absorber can look like Fig. 3.30 at right; in the same figure, on the right, a force-stroke diagram is also shown for two different speeds. This last diagram refers to a simple test machine normally used to evaluate shock absorbers; this machine moves the shock

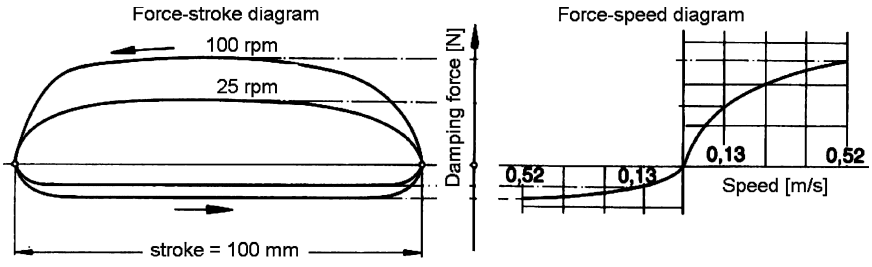


FIGURE 3.30. Diagram force-stroke of a shock absorber and conversion to a force-speed diagram.

absorber from one side with a crank, rotating at constant speed, holding it on the other through a load cell.

In this machine the stroke is constant (usually 100 mm) and the speed can be set at different values; the shock absorber extension and compression speeds are, therefore, not constant. The cycle shown in the diagram represents the extension force at the upper side and the compression force at the lower.

The diagram on the right of the figure is derived from the first, calculating for each point of the first the real compression or extension speed. The line is interpolated in the clouds of points that are obtained in this way. The extension of this cloud in the force direction is determined by the shock absorbed friction, primarily due to seals. This machine does not show the effect of side forces on the rod correctly.

The asymmetry of the diagram can be controlled by adjusting the valves.

The internal gas gives to the shock absorber a certain pre-load, obtained by multiplication of the internal pressure by the rod cross area; the polytropic law of gas justifies a variation of this force with the stroke as well.

The following figures regard other details of shock absorbers.

Figure 3.31 shows on the left the detail of the seal on a double tube shock absorber.

The seal is made by two parts: the body 1, which guides the rod 2, and the double lip seal 3. The body 1 shapes a labyrinth between the pressure and the compensation chambers.

The oil spilling from the labyrinth is conveyed to the bowl in the body 1 and does not leak to the outside because of the elastic double lip ring; this seal is not highly stressed and the lip elasticity can be reduced to consequently reduce the friction force.

The oil reservoir in the bowl can take several days do empty into the lower chamber, because of the unavoidable leakage from valves V_3 and V_4; consequently the danger of having air bubbles in the pressure chamber is reduced. Air

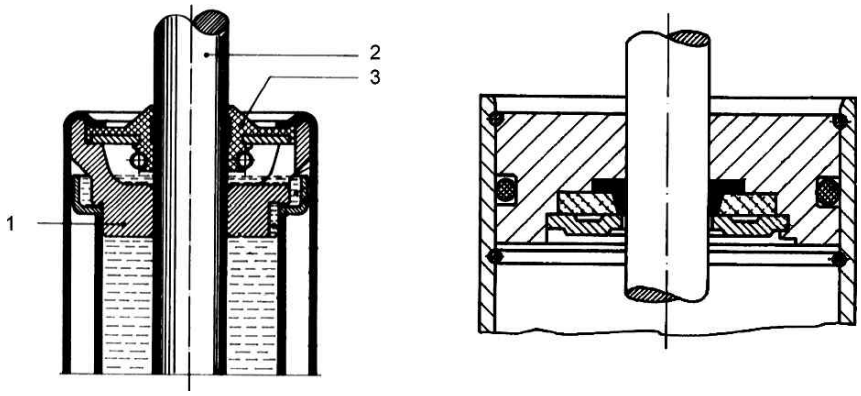


FIGURE 3.31. Seals on a double tube (left) and a mono tube (right) shock absorbers.

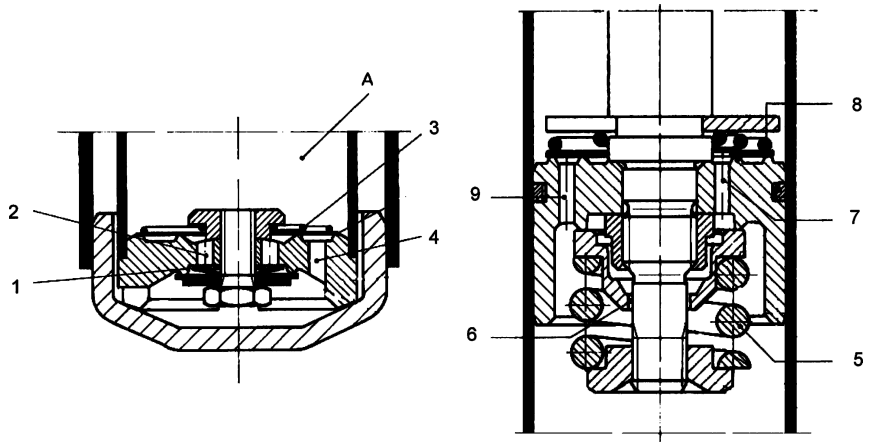


FIGURE 3.32. Valve assemblies of a double tube shock absorber.

bubbles can introduce clicking during the first suspension strokes after a long period of inactivity.

The same figure on the right shows the seal of a mono tube shock absorber.

The seal is in this case single and is formed by a ring of particular shape made of elastomers. This seal surrounds the rod tightly and effectively but hysteresis is significantly increased.

Figure 3.32 shows two different valve assemblies: One on the bottom (left) and one on the piston (right) for a double tube shock absorber. In a mono tube the differences are not substantial.

The valve on the bottom works during the compression stroke. The check valve features a bevel disc spring, on the lower side, which opens the holes 2 cut on the bottom; spring rigidity and hole diameters control the damping coefficient at low and high speed respectively.

The valve 3, on the upper part of the bottom has a flexible spring and hole 4 is quite large. The necessary pressure for oil return must be, in fact, low; this valve works during the extension stroke, to return to the A chamber the oil necessary to replace the rod volume.

The valve on the piston (at right) is controlled during the extension stroke by the coil spring 5, which allows the lift of the disc 6, opening holes 7 cut in the piston; the spring controls the low speed force, while the high speed value is controlled by the diameter of the holes.

The upper valve, closed by spring 8, works during oil transfer only, through holes 9, to compensate for the oil moved by the rod.

Springs and valve bodies are easily interchangeable, allowing shock absorbers to be built for different applications on the same production line; during the final stages of new car development the choice of dimensions for valve assemblies is usually made empirically, according to a trial and error process, starting from first approximation values obtained by mathematical models.

The characteristic curve of the damping force versus speed can be classified, with reference to its shape, as:

- *Constant*, when the derivative of force with respect to extension or compression speed is almost constant, but usually with different values in compression and extension
- *Progressive*, when the derivative increases with speed
- *Degressive*, the opposite case, when the derivative decreases with speed

The derivative of the force with respect to speed is also called the *damping coefficient*; a constant damping coefficient appears ideal according to the Boursier De Carbon theory, as we will explain in the second volume; a progressive characteristic could be suitable to improve the handling quality of a sports car, with some sacrifice in comfort; a degressive characteristic can, instead, improve comfort on bumpy roads, where suspension stroke speed can be high.

The rubber bushing between shock absorber and car body is used to filter dynamic forces over $30 \div 40$ Hz of frequency, where shock absorbers looses their capacity significantly; these vibrations cause noise that diffuses to the car interior through the body structure.

In McPherson suspensions, the bushing integrated within the upper pivot can be designed in two ways: Simple or double.

In the simple configuration the shock absorber loads the body together with the coil spring and the stop spring; if the load is significant rubber must be well sized and is usually stiff.

In the double configuration, a dedicated rubber bushing carries the shock absorber load only. Static load is negligible, except for the gas pressure effect, if any. In this case the rubber can be lighter, with improvements in flexibility and a better filtering capacity.

For non-structural shock absorbers the body mount is made by a simple cylindrical rubber bushing (detail a of Fig. 3.33); the symmetry axis of the bushings must be oriented to make the inclination changes due to the suspension stroke easier.

If the value of this change of inclination is contained ($<$10 deg), the inside and outside cylindrical surfaces of the rubber bushings are press fitted between the hole and the pin, with the advantage of avoiding lubrication; in this case the shear of the bushing stresses the piston and the piston rod of the shock absorber.

If the angle is greater, it is better practice to use a self-lubricating additional sliding bushing to reduce hysteresis.

If other rotation angles are present it is better to shape the rubber bushing as in detail c.

For very limited rotation angles ($<$5 deg), the simple mount shown in detail d can be adopted.

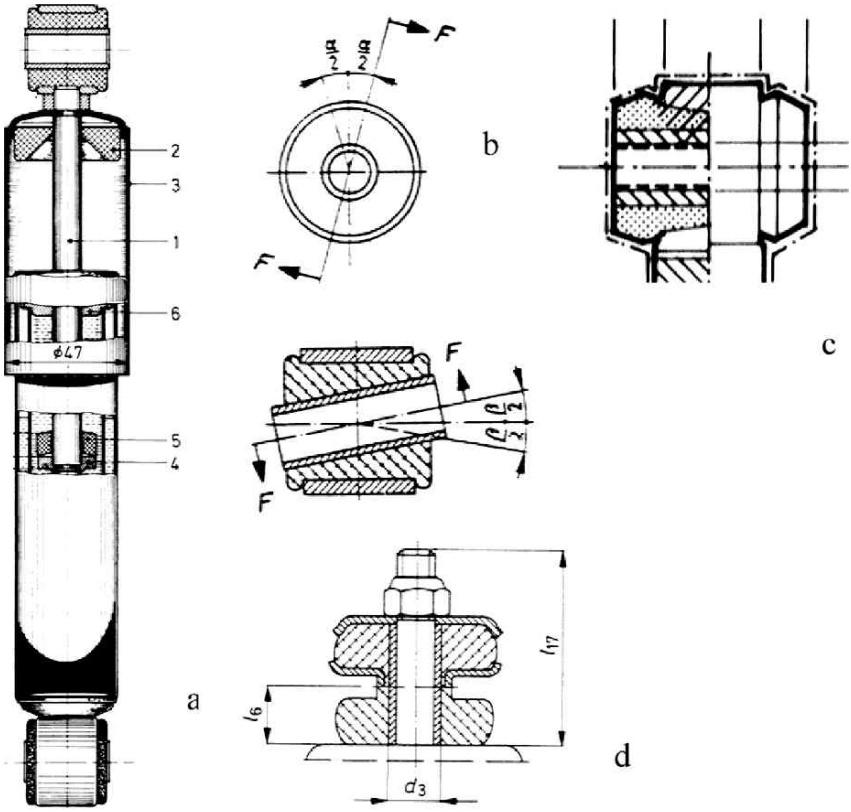


FIGURE 3.33. Kinds of body and suspension mounts for a non-structural shock absorber; the conical stiffness of these mounts can affect shock absorber friction.

3.2.2 McPherson suspensions for rear axle

As we have previously stated McPherson suspensions can be applied to rear axles as well.

In general, a rear axle suspension must perform the following functions:

- Must be compatible with the installation of many bulky components, such as fuel tank, spare wheel, muffler and exhaust pipe
- For rear wheel drive, must provide the space for differential and final drive and transmission shafts
- Must allow the body a space for as large a cargo compartment as possible
- Must allow a suspension stroke longer than for the front axle, because rear axle load variations are larger

Fortunately the wheel position envelope is smaller because steering angles are limited to the much smaller toe angle variations.

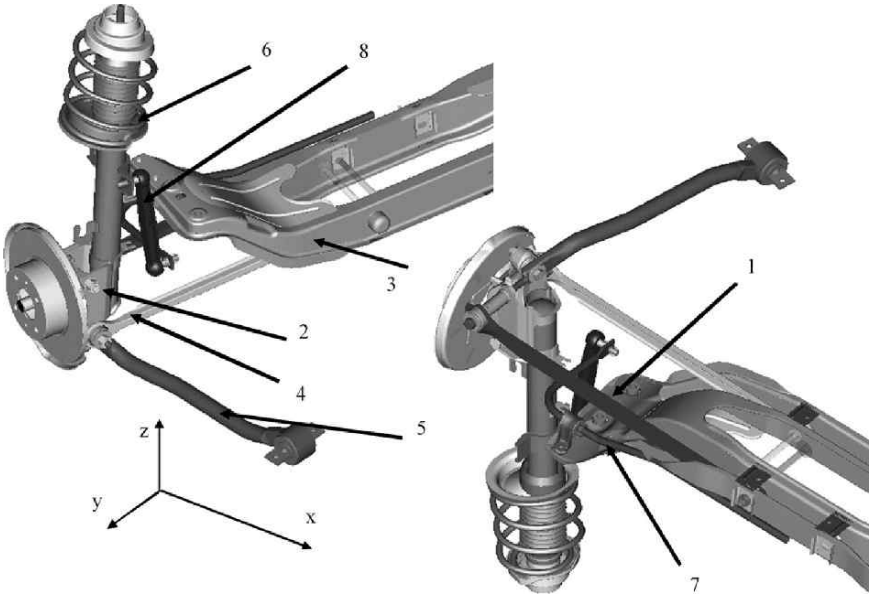


FIGURE 3.34. Rear McPherson suspension for a medium size car with front wheel drive. Can be derived from the same type of suspension for the front axle, by exchanging the steering rod with the lever 4 (Alfa Romeo).

The McPherson configuration allows better elasto-kinematic performance when compared with the other typical solution for rear trains we will see later, with some sacrifice in cargo compartment width and component cost.

Figure 3.34 shows the rear suspension of a medium size passenger car.

The kinematic scheme of this suspension can be derived from that of the front axle suspension by changing the steering link with a bar 1 connected between the strut 2 and the subframe 3. The cross bar 4 is for cornering forces, while the bar 5 is for longitudinal forces; together they work as the lower arm of the front axle suspension.

The application of a subframe does not affect in first approximation the elasto-kinematics of the suspension.

The spring-shock absorber unit 6 is articulated to the body to a dome in the wheel case. The anti-roll bar 7 is supported by the subframe and is linked to the shock absorber with a pendulum rod 8.

The appearance is quite different from that of the front wheel suspension even if the kinematic scheme is similar and the number of links is the same.

The three links 1, 4 and 5 lie on the same plane; links 4 and 5 have the same function of an arm with virtual center; in the same way link 1 is like a steering bar.

Either because of the mutual position with the wheel or the higher stiffness of its bushings, the bar 4 will absorb a high percentage of cornering forces.

Because of its orientation, the bar 5 will absorb a high percentage of the braking forces and other longitudinal loads; its bushing connecting with the body should be as flexible as necessary for a better filtration of the longitudinal shocks caused by obstacles, without affecting toe angles, because the steering mechanism is again an articulated parallelogram.

Finally, the flexibility of the bushing of the articulation with the body of bar 1 can be finalized to correct the understeering behavior of the car under the action of cornering forces. In conclusion, each of the elastic bushings can be designed for a single performance, without negative consequences on the remaining bushings.

Examples of rear McPherson suspensions also exist that are more similar to the front axle type; in this case, the strut rotation causing wheel steering is prevented by a steering bar fixed, at the inner end, to the body. This is usually called a *false steering bar*; the articulation position of the fixed end can be designed to obtained the desired toe angle variation.

The advantages and disadvantages of rear axle McPherson suspension can be summarized in the following list.

Advantages

- Good camber recovery with suspension stroke.
- Potential for proper toe angle variation by cornering forces.
- Potential for proper longitudinal flexibility of the wheel.
- Reduced unsprung mass.
- Suitable for driving axles.

Disadvantages

- Forces acting on the body through the dome are applied to a flexible point.
- The coil spring-shock absorber unit reduces the cargo compartment width.
- A structural shock absorber offers potential hysteresis problems.
- Medium complexity and cost.

Because of these characteristics the rear axle McPherson suspension is used on medium size luxury cars or on sport cars.

3.2.3 Double wishbone suspension

Double wishbone suspensions are applied to luxury sedans and sports cars because they allow a design of the elasto-kinematic parameters that provides an optimum compromise between handling and comfort.

Because the upper arm is in general shorter than the lower to allow a given camber recovery, these are also called short and long arm suspensions, with the acronym of SLA suspensions.

Description

Figure 3.35 shows two different kinds of double wishbone suspension; the first on the left, is suitable for front wheel drive cars with transversal power train; the second, on the right, can be applied to longitudinal power trains.

If we compare these with the McPherson suspensions, we note the upper arm, which carries on the function of the shock absorber sliding bearing. The shock absorber has no structural functions because the damping force – not static and dynamic forces – is what loads it.

Hysteresis and the consequent penalty on comfort is therefore limited.

According to the upper arm position with respect to the wheel, suspensions of this type are classified as high (on the left) or low (on the right).

The difference between these two solutions is imposed by the transversal bulk of the engine and the track that must be obtained for the car; as a matter of fact, the high double wishbone suspension is the only one that can be adopted on front wheel drive cars with transversal engines. It is applied to other cars also, because it can suit the structure of a unitized body well. Low double wishbone suspensions, featuring excellent elasto-kinematic performance, intrude into the engine compartment and are generally applied only to luxury touring cars.

Figure 3.36 shows a high double wishbone suspension in detail; we refer to what we have said about the McPherson suspension for many common details, such as steering bars, subframe, steering mechanism, anti-roll bar and elastic bushings.

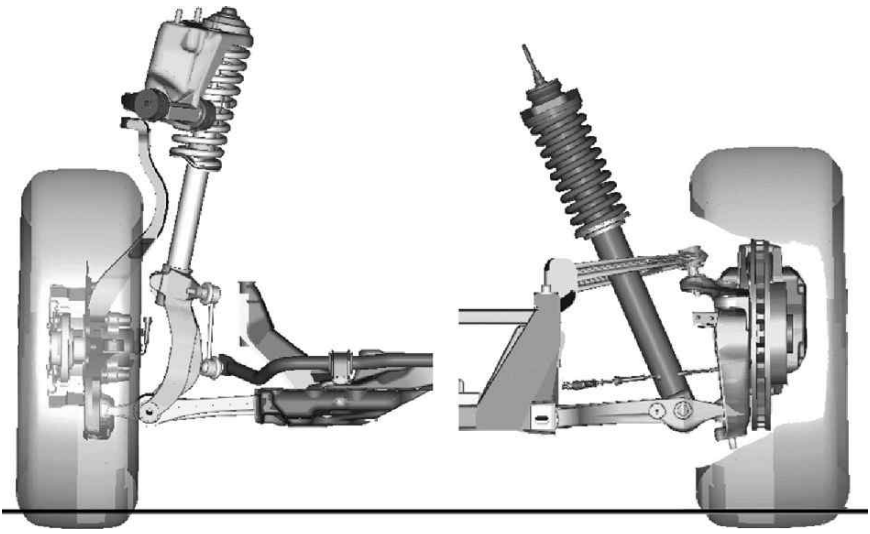


FIGURE 3.35. Front wheel double wishbone suspensions, of the high type, on the left and of the low type, on the right; only the high double wishbone suspension can be applied to transversal power trains.

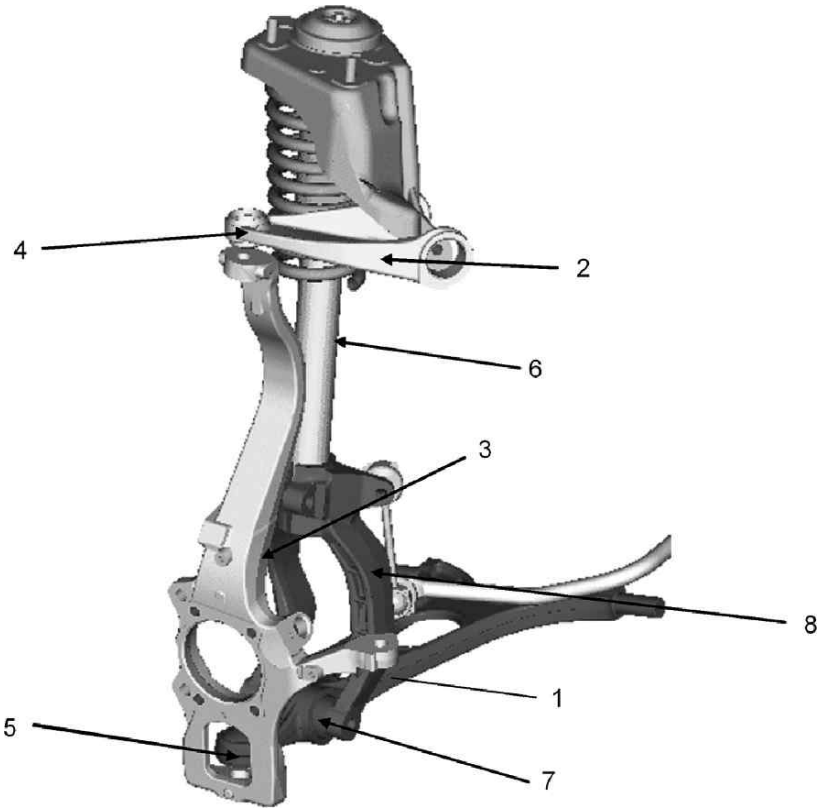


FIGURE 3.36. The lower arm 1 and the upper 2 are linked to the car body with rubber bushings. These arms are also linked, at the other end, to the strut 3, through spherical joints 4 and 5. The fork arm 8 connects the lower arm with the shock absorber (Alfa Romeo).

The upper arm 1 and the lower 2 are linked to the body through elastic bushings. The same arms are linked to the strut 3 through spherical joints 4 and 5 which allow the steering rotation of the strut. The line connecting the two spherical joints is the king-pin axis.

The shock absorber and the coil spring are unitized as in the McPherson suspension; this assembly is connected to the lower arm through the elastic bushings 7 and is shaped like a fork, to clear the bulk of the half shaft.

The shock absorber piston rod is connected to the body through an elastic bearing; in this case there is no pivot bearing, because the steering motion involves the wheel strut only. A rest includes upper arm bushings and the shock absorber bearing and is bolted to the body.

In other applications this rest can have a different shape or can be eliminated, with bushings and bearings linked directly to the car body; this solution

is justified by the need for standardization. The body can be produced in the same plant together with other bodies equipped with the McPherson suspension.

The lower arm is connected to a subframe. The torsion bar is connected to the shock absorber through a pendulum rod. As in previous suspensions, the subframe is used for the steering rack installation as well.

Advantages and disadvantages

Advantages

- Optimum design of elasto-kinematic parameters, particularly as far as camber recovery is concerned.
- Shock absorbers have no structural function; comfort can be improved, because of hysteresis reduction.
- Possibility of lowering the hood profile, particularly for the low version.

Disadvantages

- Production costs are higher, because of the increased number of parts; an arm and its bushings are added, as compared with the McPherson suspension.
- Additional parts for upper suspension attachment.
- The space taken by the upper arm is notable. Transversal engines demand the high version: The reduced length of the upper arm of this version compromises the possibility of reaching the maximum elasto-kinematic performance.
- The increased number of joints and bearings can affect wheel angles because of permanent deformations in the bushing rubber, with negative consequences on tire wear.
- The high value of braking loads can have a negative influence on longitudinal flexibility.

Design details

In high wishbone solutions the strut is characterized by a gooseneck shape, if observed on an yz plane. This shape is conditioned by the bulk of the tire and the upper joint position, with reference to the engine. This thin and highly stressed component is usually made from hot stamped steel.

The example in Fig. 3.36 shows an arm inclination in the xy view; this configuration is demanded by the transversal position of the power train.

Figure 3.37 shows an alternative design of a low double wishbone suspension, with reference to spring and shock absorber position.

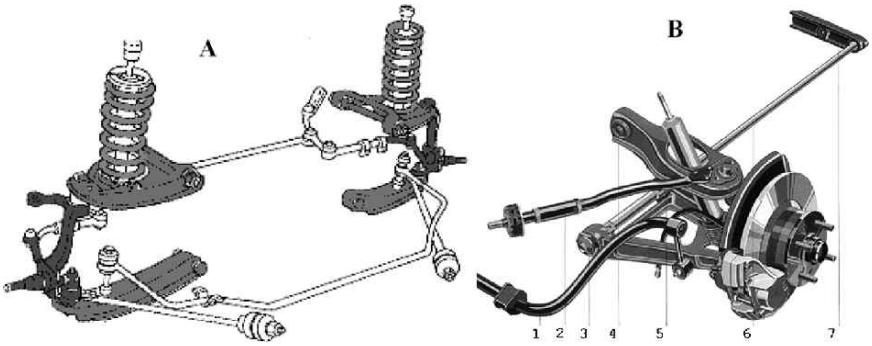


FIGURE 3.37. Different alternative positions for the coil spring and the shock absorber for a low double wishbone suspension.

Solution A is characterized by shock absorbers and spring being applied to the upper arm.

Solution B shows a spring made by a torsion bar, with the shock absorber applied to the lower arm: This solution allows a limited hood height and introduces limited values of stress in the upper body parts. The same result, in terms of height alone, can be obtained by using a unitized coil shock absorber, assuming there is sufficient space.

A final variant of this suspension can have, as the springing element, a transversal leaf spring, which can incorporate the function of one of the arms; the distance between the bearings of the leaf spring can be designed so as to obtain the desired symmetrical and asymmetrical flexibility without using an anti-roll bar. This solution is rarely applied today because of the hysteresis penalties of the leaf spring.

3.2.4 Virtual centres suspensions

In the suspensions we have examined to this point, the strut is linked to two spherical joints, which identify king-pin axis. In many cases the king-pin axis cannot be set in the position needed to obtain a desired value (limited or negative) of the king-pin offset, without putting an undesired camber angle on the wheel.

Some manufactures solve this problem by adopting, particularly in high end cars, single or double *virtual center* links for the suspension strut; here one or two arms are replaced by a doubled number of linkages, each with its spherical joint. The king-pin axis is no longer identified by the physical position of a joint, but by a virtual point given by the intersection of the two lines connecting the articulation points of a linkage.

This scheme can be applied to both arms or the lower arm only, which is critical for obtaining a negative king-pin offset. In this second case the lower arm of a McPherson suspension can also receive a virtual centre articulation.

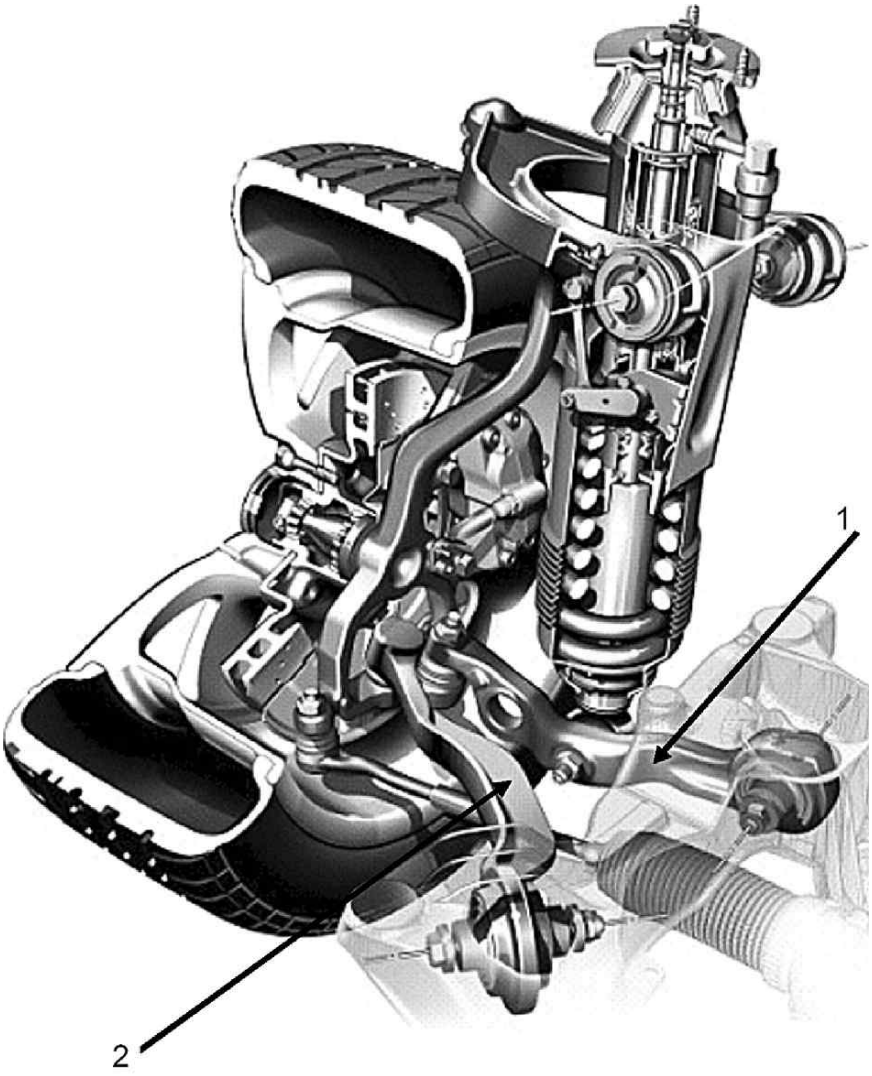


FIGURE 3.38. High double wishbone suspension with lower virtual center articulation; the king-pin axis is determined by the intersection of the two lines through the articulations of arms 1 and 2 (Mercedes).

In cases where the lower arm alone adopts virtual centers, this last will be replaced by two arms, as in Fig. 3.38, while the upper arm will be shaped conventionally.

The king-pin axis can be determined by the already established geometric construction, by intersecting the two planes through the lines connecting the four articulation points of the arms and the upper physical center, as shown in Fig. 3.38.

In McPherson suspensions the upper physical center is the upper pivot.

The virtual center configuration also allows a negative king-pin offset for large tires and bulky brake discs. It is known that a negative king-pin offset can improve vehicle stability during braking, when one of the braking circuits has failed or when different friction coefficients are applied to the two wheels of the front axle.

The disadvantage of this solution is the introduction of additional spherical joints. Arms must be shaped in a particular way so as to avoid the wheel envelope space.

With this design feature the king-pin axis changes with the steering angle; as a consequence the king-pin offset must be verified for each steering angle.

As outlined, this solution can be applied as well to the upper arm of a double wishbone suspension, as shown in Fig. 3.39. As we have explained in the previous case, the king-pin axis can be obtained by intersecting the two planes through the two lines for the four articulation points of the two rear arms and the two front arms.

It is possible to reduce the value of the king-pin offset even for large disc brakes.

King-pin offset reduction can also have benefits on steering wheel return, which can be affected by the value of the traction force.

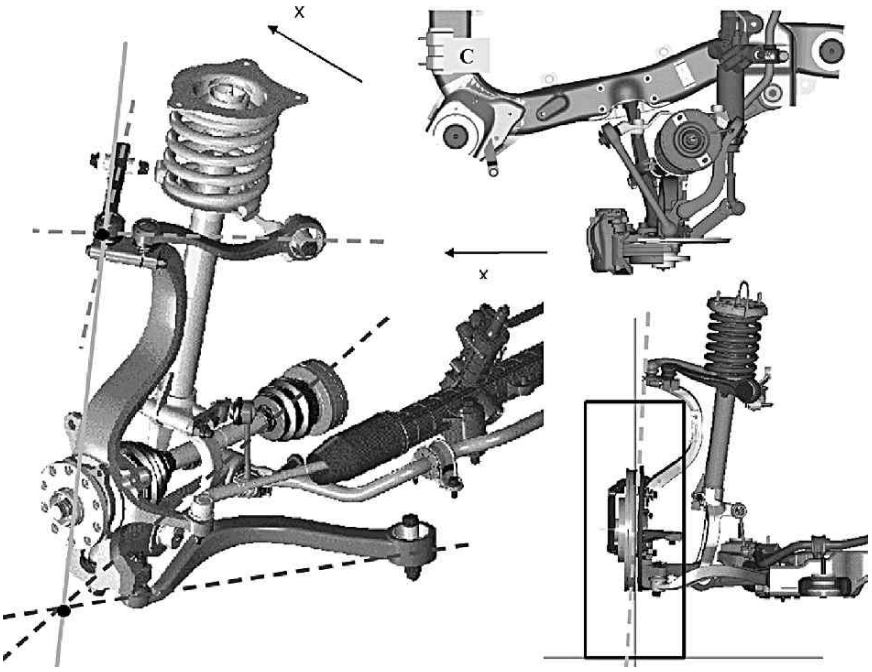


FIGURE 3.39. High double wishbone suspension with both virtual center suspension arms; the figure shows the geometrical construction needed to obtain king pin axis position (Lancia).

3.2.5 Trailing arm suspensions

This architecture is widely applied in small and medium cars on rear axles only.

Both wheels are fixed to a longitudinal arm free to rotate with reference to the body: The rotation axis of the two arms is usually the same and is parallel to the y vehicle axis.

During suspension strokes the wheelbase is affected but toe angles remain unchanged. Wheel camber with reference to the ground is equal to the car body roll; there is no camber recovery.

Description

Figure 3.40 shows two perspective views of a trailing arm suspension for a front wheel driven medium class car.

The axle features two oscillating arms connected to a subframe through a transversal rotation axis; the subframe is connected to the body by four elastic bearings. This subframe is made from two shells of stamped steel sheet, welded to a tubular cross beam. If subframe is not applied, arm bearings can be fixed directly to the body.

In each shell is set the upper spring rest. A bracket offers the upper shock absorber mount.

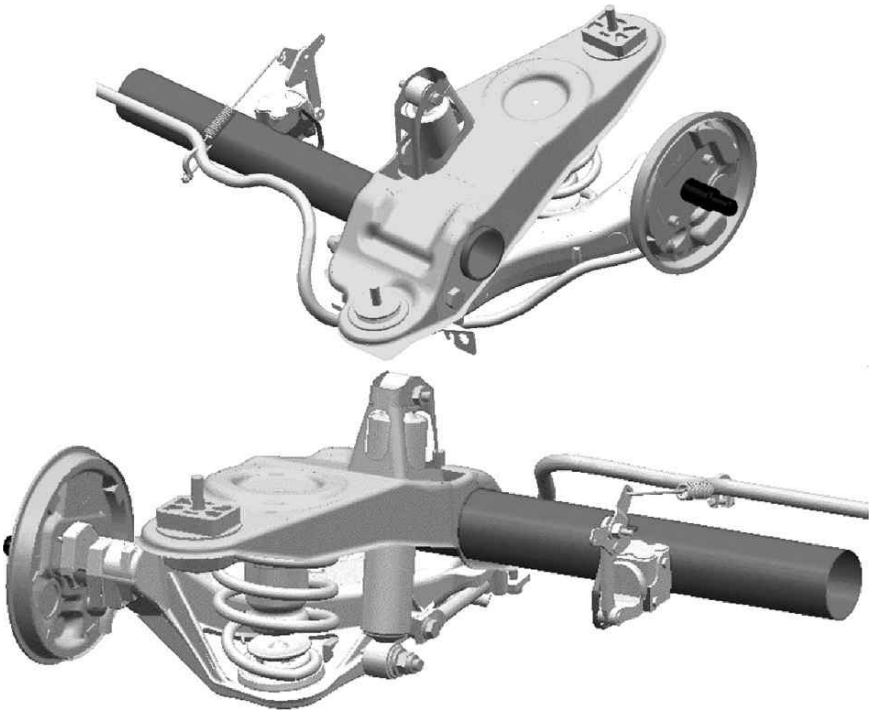


FIGURE 3.40. Rear suspension with trailing arms (FIAT).

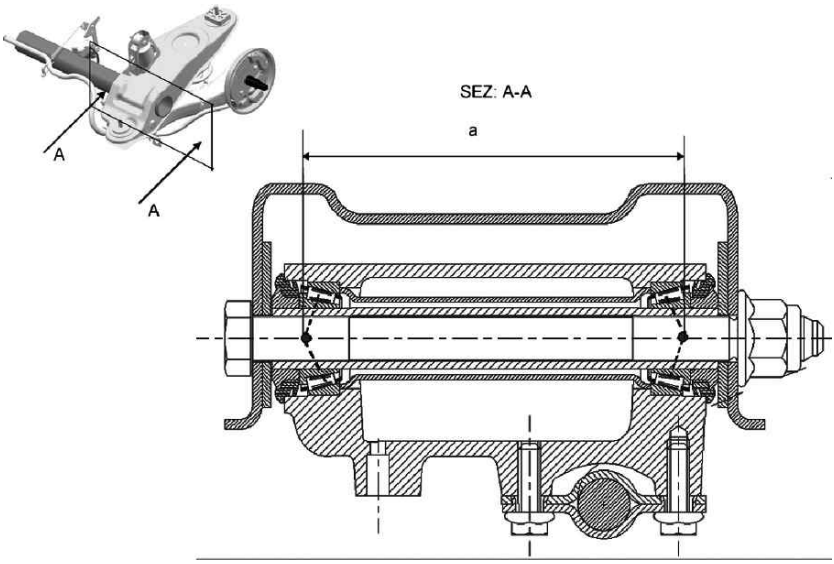


FIGURE 3.41. Detail of tapered roll bearings of the arm of the trailing arm suspension of the previous figure; one of the fixation points of the floating anti-roll bar can be also seen.

The suspension arm is here made from an iron casting, integrating the lower, moving coil spring rest, shock absorber bushing case and wheel bearing case or pin.

The anti-roll bar is fixed to the arms directly at two different points on each arm; there are no fixation points on the body. For this reason it is also called a *floating* bar.

Figure 3.41 shows a detail of the arm of the above suspension. It shows an opposed conicity roller bearing used for the articulation of the arm to the subframe. Because this bearing is quite stiff, there will be no toe angle, or camber variation, or other displacements caused by arm deformations because of forces applied to the wheel.

In terms of the camber angle we must also take into account the arm torsional deformations caused by vertical and lateral forces.

The choice of roller bearings instead of the more traditional rubber bushings is demanded by the high value of the reaction forces, due to the reduced distance a between the bearings; a higher value could allow the use of rubber bushing but would reduce the space available for the fuel reservoir and spare wheel.

Figure 3.42 shows a cross section of the wheel assembly. The wheel bearing inside ring is press fitted to a pin, which is again press fitted into the suspension arm; the bearing is second generation. The arm shows a flange for fixing a drum brake or the caliper of a disc brake; notice that the braking torque applies to the arms a reaction torque, which tends to reduce the body rear up-lift due to the inertia force.

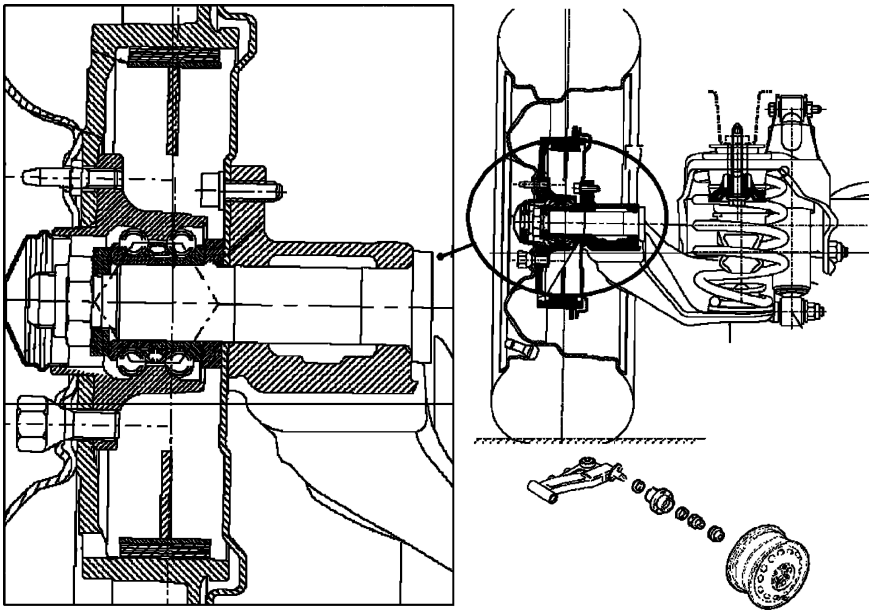


FIGURE 3.42. Detail of the fixed wheel pin of the second generation bearing of the wheel hub of a trailing arm suspension.

This example applies two coil springs as elastic elements; torsion bars can also be applied to reduce the vertical height of the cargo floor.

In this case, three torsion bars are customarily applied: One for each suspension is dedicated to symmetrical springing and fixed at the arm on one side and the subframe on the other. A third bar is fixed at each side on the two suspension arms and acts as an anti-roll bar in asymmetric springing. Sometimes this kind of arrangement causes different wheelbases to be used on each side of the car.

In the past this kind of suspension has also been applied to steering front axles, as in Fig. 3.43, but this solution is no longer used.

In fact king-pin axis direction changes either by symmetric or asymmetric strokes and the return torque of the steering wheel can increase excessively on curves, because of the increase of the longitudinal trail of the external, more loaded wheel. In addition, the longitudinal motion of the wheel, because of vertical stroke, is opposite to the direction along which obstacles are encountered, with negative results for comfort.

Advantages and disadvantages

Advantages

- Hysteresis is very low with roller bearing versions.
- Suspension intrusion into the baggage compartment is minimal.

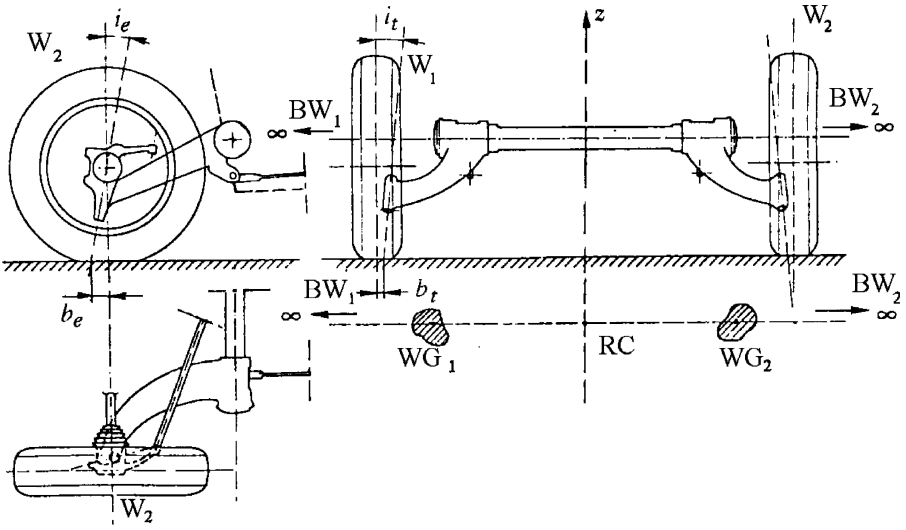


FIGURE 3.43. Front axle with longitudinal trailing arms; springs working through pull bars are not represented. Roll center position has been calculated taking tire deformation into account. Different rotation centers are represented (see the dedicated paragraph on this topic).

- High simplicity and, therefore, low production cost.
- Ease of assembly.
- Suitable for driving axles as well.
- Reduced value for the unsprung mass.

Disadvantages

- Transversal deformations of trailing arms caused by cornering forces have an oversteering effect.
- No camber recovery.
- Low longitudinal flexibility because of the rigidity of highly loaded bearings.
- There are no independent parameters to be tuned for improving the elastokinematic behavior.
- High vibration transmittance from the wheel because of bearing stiffness, again due to the value of acting loads.

This solution is used predominantly in low-end market segments.

3.2.6 Semi-trailing arms suspension

As a difference with trailing arm suspensions, which feature arms rotating around the same transversal axis, semi-trailing arms or triangles rotate around two different symmetric axes, with reference to the car zx symmetry plane.

This axis shows an inclination angle either on the zy plane or the xy plane; a particular case of this kind of suspension is the cross arm suspension, where the arms rotate around a longitudinal axis.

This suspension is applied to rear axles only.

Figure 3.44 shows a comparison between a trailing arm scheme (a) and a semi-trailing arm scheme (b).

Arm inclination angles allow a modest camber recovery to be obtained and a certain toe angle variation with understeering effect, with a slight improvement of the elasto-kinematic behavior.

Description

Figure 3.45 shows the semi-trailing rear suspension of a small front wheel driven car.

The axle is formed from two swinging triangular arms (1 on scheme b) articulated to a subframe (1 on scheme a), being fixed to the car body at points A and B.

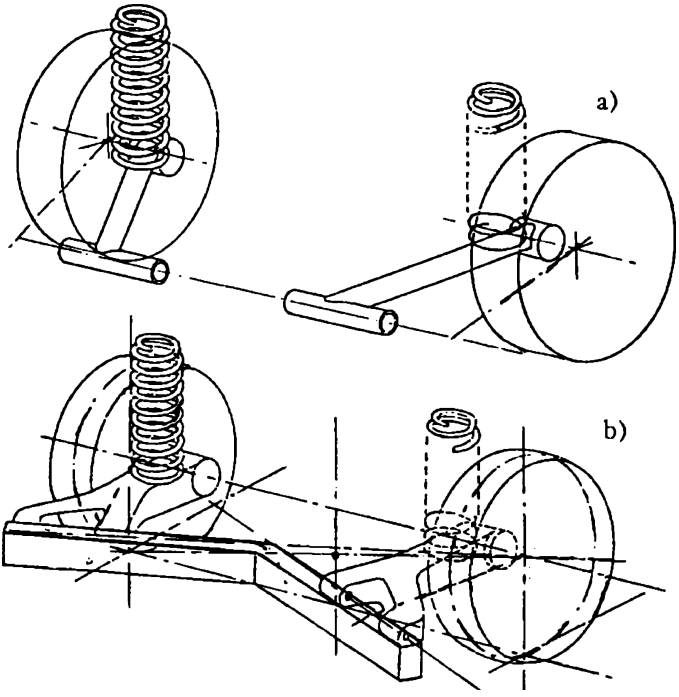


FIGURE 3.44. Trailing arm suspension (a) and semi-trailing arm suspension (b).

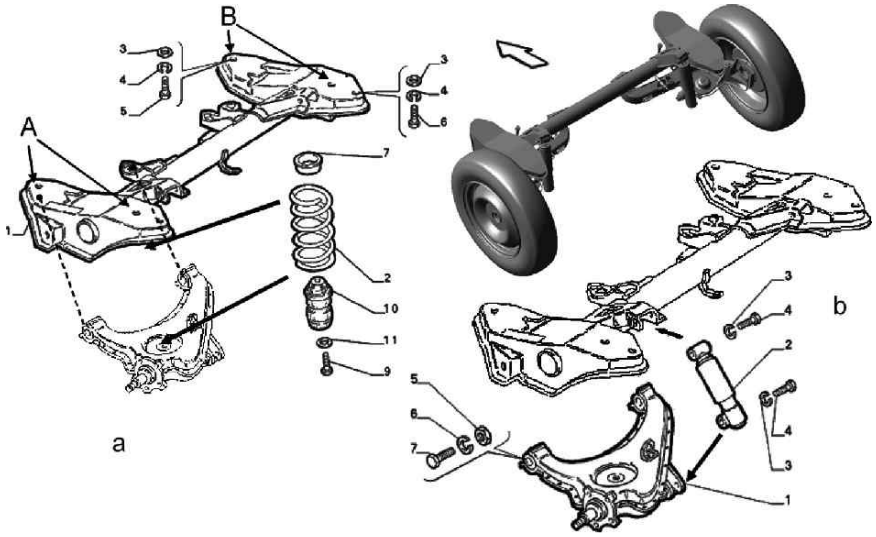


FIGURE 3.45. Assembly view and details of a semi-trailing arm suspension for a small front wheel driven car (FIAT).

The subframe is made from a cylindrical tube with two stamped steel shells at the two ends; the arms are also made from two stamped semi-shells spot welded together. This figure also shows the position of the coil spring and the stop spring (2 and 10 in scheme a) and of the shock absorber (2 in scheme b).

Unlike trailing arm suspensions it is here possible to use elastic bushings for arm connections to the body, thanks to the increased distance between the two articulation points.

We can say that this distance is not caused by the type of suspension but rather by the dimension and installation of the underbody components; in this car the increased space is justified by the installation of the spare wheel under the hood, possible because of its small diameter.

Although this application has only historic value because of its extreme roughness, we show in Fig. 3.46 a semi-trailing arm suspension suitable for rear engine cars; this example also explains the main reason for using this kind of suspension, that it simplifies the design of drive shafts.

The arm rotation axis can cross the center of the constant speed joint; this condition avoids shaft length variation and allows the application of very simple joints.

Because of the inclination of the rotation axis with reference to the body symmetry plane a partial camber recovery can be achieved. This practice is limited by the accompanying toe angle variation that, within certain limits, can improve the natural oversteering of a rear engine car but may not be viable over certain values.

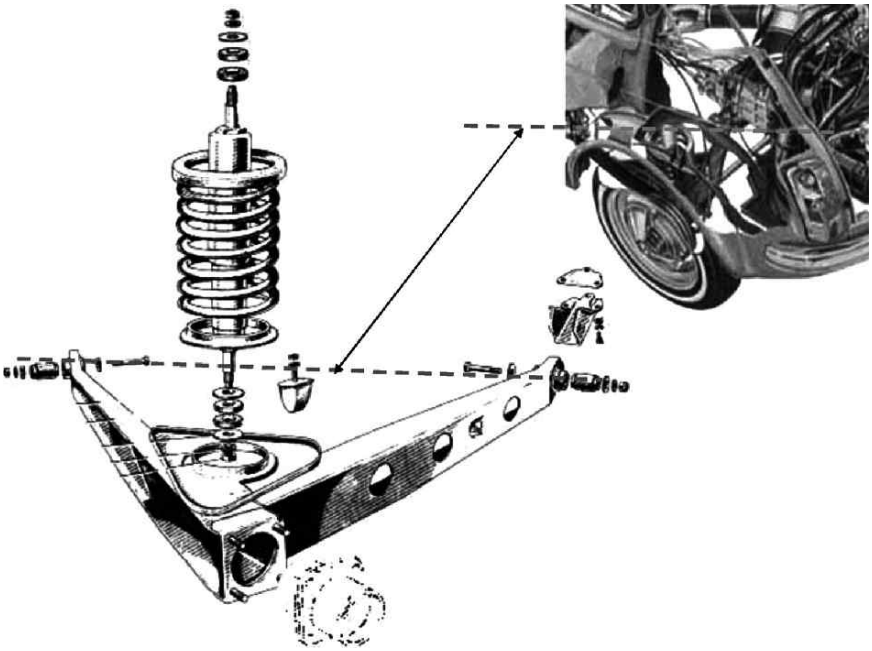


FIGURE 3.46. Semi-trailing arm suspension in a small car with rear wheel drive and rear engine; the arm rotation axis crosses the center of the constant speed joint (FIAT).

Advantages and disadvantages

Advantages

- Limited vertical dimension
- Limited unsprung mass
- Reasonable design possibility in terms of elasto-kinematic properties
- Construction simplicity
- Suitable for rear wheel drive

Disadvantages

- Transversal dimension penalizes underbody component layout
- Excessive track variation because of the suspension stroke can wear tires prematurely

This solution is presently used only on microcars or quadricycles

3.2.7 Guided trailing arm suspensions

Guided trailing arm suspensions have been conceived as an improvement to trailing arm suspensions; with respect to the latter suspension more linkages have been added and the number of degrees of freedom of the trailing arm has been increased, in an attempt to conserve the natural advantage of this suspension: a reduced intrusion into the cargo compartment.

There are not, at present, names for these suspensions shared universally by chassis engineers: We have selected this one to point out that there is a more sophisticated connection of the trailing arm to the body than a simple rotary bearing.

Many car manufactures, for commercial reasons, have improperly labeled this a multilink suspension; in the opinion of the Authors this name should be reserved to the five linkage suspensions only.

In guided trailing arm suspensions, two or three additional arms are linked to the trailing arm, in order to improve the elasto-kinematic performance of the suspension.

To restore the correct number of degrees of freedom of the mechanism, the connection between trailing arm and body or subframe is made by a sufficiently flexible rubber bushing.

Figure 3.47 shows an example of a guided trailing arms suspension; the trailing arm 1 is guided, in this case, by two additional cross arms 2 and 3.

Cross arms identify a steering axis through the two elastic bushings 4 and 5. This axis is designed to create a given understeering toe angle variation, under the

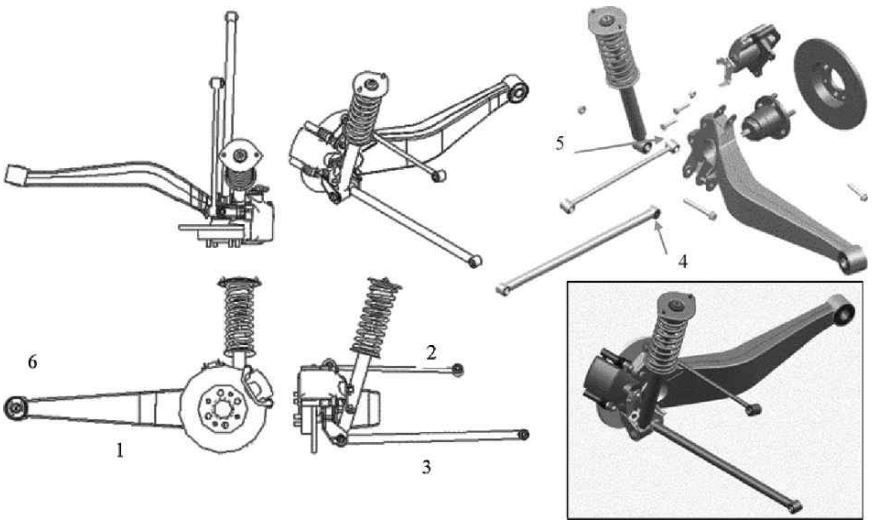


FIGURE 3.47. Scheme for a guided trailing arm suspension; the articulation point 6 must allow two rotational degrees of freedom to feature the desired camber angle recovery.

action of the cornering force or braking force. The elastic bushing 6 of the trailing arm, as the dimensions suggest, can provide excellent longitudinal flexibility and thus enhanced comfort.

The position of bushings allows longitudinal motions without undesirable steering rotation of the wheel.

Figure 3.48 shows a different example. The application of three cross arms can be noticed in this suspension. The introduction of the third arm further improves the elasto-kinematic behavior.

With reference to the lower figure, arms 1 and 2 shape a transversely articulated parallelogram, which can provide the correct camber recovery. The linkage 2 is made by a shell structure of stamped steel sheet, relatively stiff because spring loads are applied to it; linkage 1 is, on the contrary very light, because only transversal forces are applied to it. This linkage is curved to fit around the side beam of the underbody.

The linkage 3 can react to longitudinal loads with controlled deformation to stabilize steering with an understeering effect, as we have seen in other suspensions.

The trailing arm 4 is also made by steel sheet and shows a notable flexibility; if it were rigid there would be too few degrees of freedom and the suspension would be blocked. The arm can react only to forces acting through the articulation points in the longitudinal direction, and to bending moments such as those supplied by the braking force; camber variations can occur because of its torsional deformations. With a suitable elastic bushing the wheel can move longitudinally on obstacles, with benefits for comfort, without penalty on the toe angle variation.

Transversal arms are connected to the subframe 5, suspended from the body through elastic rubber bearings.

Advantages and disadvantages

Advantages

- Toe angle variation under cornering and longitudinal forces with stabilizing effect
- Suitable camber recovery
- Good longitudinal flexibility
- Limited under floor volume similar to simpler suspensions

Disadvantages

- Wheel case restriction if shock absorbers are applied to trailing arms
- Many adjustment points for correct assembly on the subframe
- Higher cost complexity as compared to previous solutions

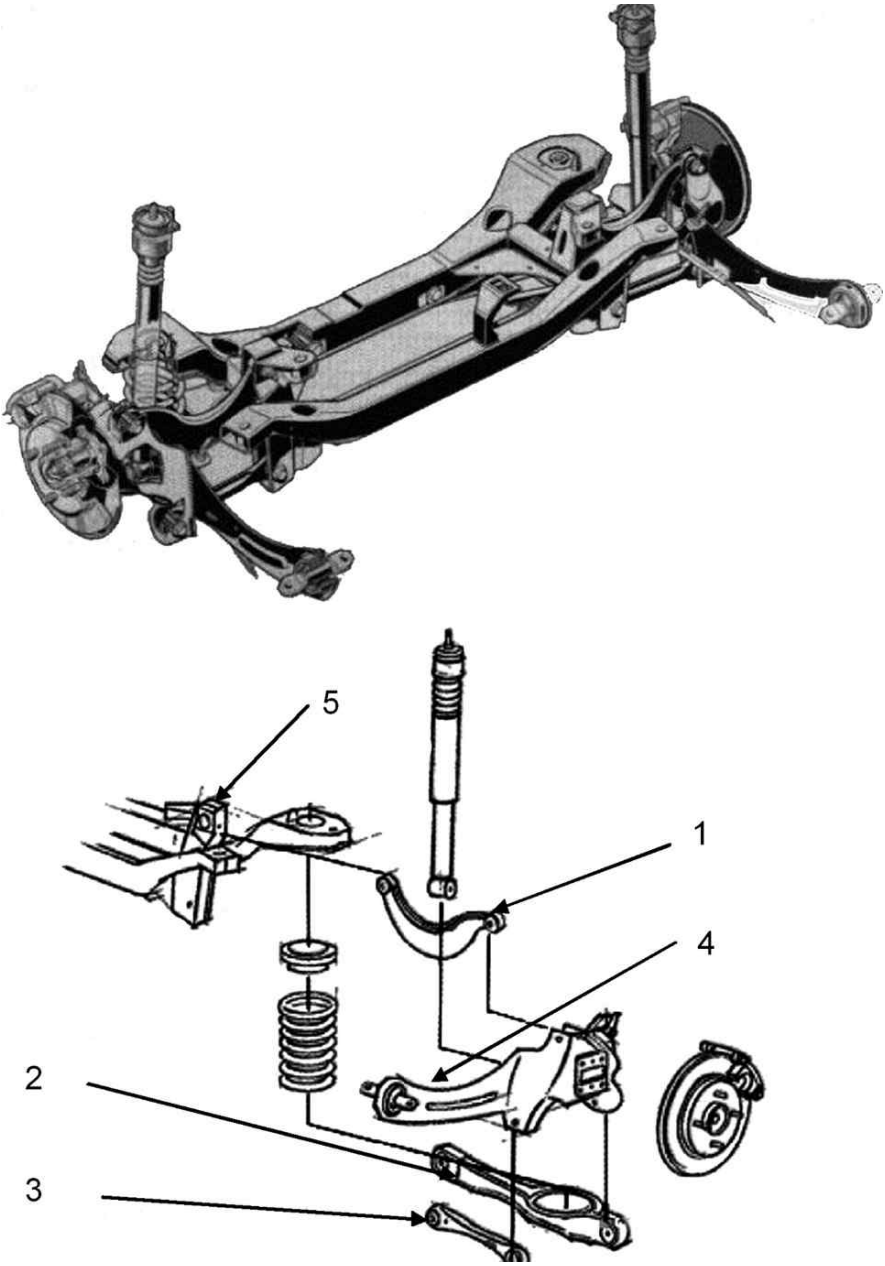


FIGURE 3.48. Outline drawing of a guided trailing arm suspension; to understand its operation the torsional deformation of arm 4 must be taken into account (Ford).

Because of these characteristics guided trailing arm suspensions are used on medium cars, where it represents a good compromise between sophisticated multilink suspensions and lighter trailing arm or McPherson suspensions; this architecture could be widely adopted in the future.

3.2.8 Multilink suspensions

The solution shown in Fig. 3.49 obtains, with penalties on weight and cost, the best result for comfort and handling.

In multilink suspensions the strut is connected to the body with five linkages, as many as the degrees of freedom to be subtracted from the strut, leaving only the suspension stroke motion.

This or similar schemes are applied on most large and luxury car rear axles, for both front and rear wheel drive.

This suspension can also be considered as a double wishbone suspension with both lower and upper steering virtual points (4 and 5 for the upper arm and 2 and 3 for the lower). The fifth linkage, or false steering linkage, is used to control the wheel steering motion.

The difference from a front wheel double wishbone suspension is limited; nevertheless elasto-kinematic behavior will be different considering the different role played by the two axles.

Every fixed articulation point is set on a subframe because of the previously explained benefits for pre-assembly capability and comfort.

When adopted on a rear driving axle, toe angle variations depend on the traction force, in addition to the cornering and braking forces; this fact must be taken into account while designing rubber bushings.

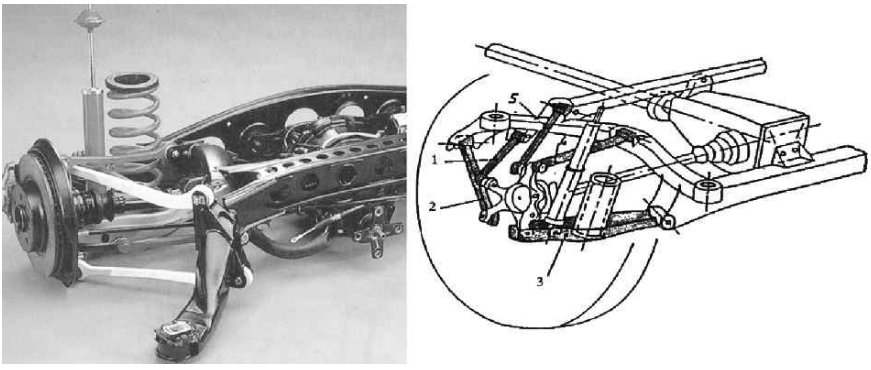


FIGURE 3.49. Classic multilink suspension for the rear axle of a rear wheel driven car, with five linkages, one for each degree of freedom subtracted from the wheel to allow the suspension stroke only (Mercedes).

Any function can be obtained in the field of:

- Toe angle variation, stabilizing, turning and braking
- Longitudinal flexibility, without undesirable toe angle variations
- Camber recovery by roll angle

In addition, for rear wheel driven cars, the wheels can increase their toe-in angle as a function of the traction force, with reduction of the phenomenon of steering angle reduction at increasing traction force.

This phenomenon, called *torque steering*, is justified if we recall that the application of a longitudinal force increases the side slip angle of the rear wheels for a given cornering force. To avoid a path change, the front wheel steering angle must be decreased.

In summary the design freedom allowed by this suspension is notably increased, particularly for a driving axle.

Similar considerations also apply to the multilink suspension of a large front wheel drive sedan (Fig. 3.50).

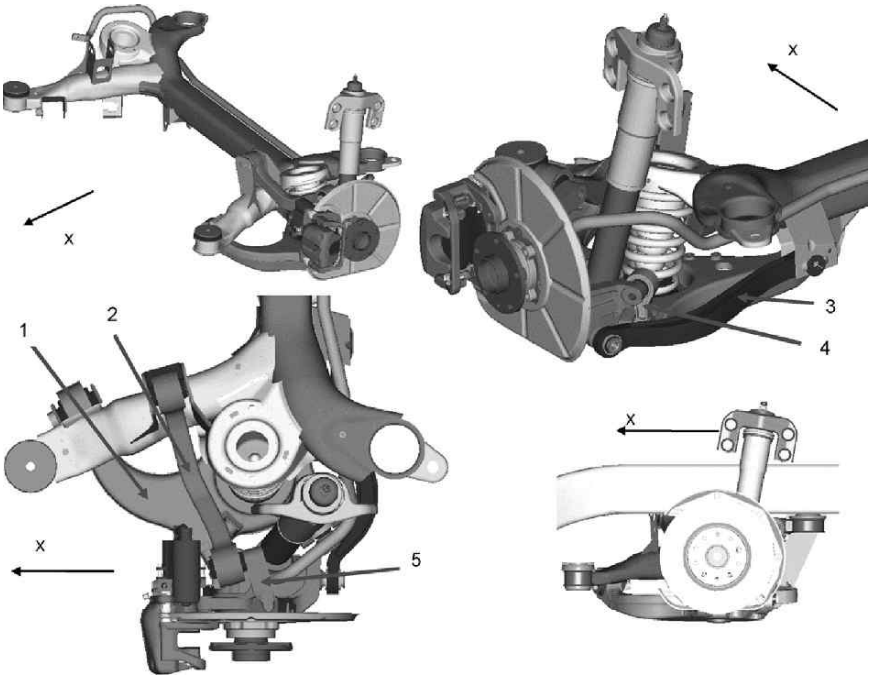


FIGURE 3.50. Particular kind of multilink suspension for front wheel driven cars; the linkage 4 is used to increase wheelbase variations during the compression stroke with benefits for comfort (Alfa Romeo).

This solution is characterized by the triangular arm 1, a cross beam 2 and a false steering linkage 3; the remaining linkage 4 connects the triangle to the suspension strut.

Once again the scheme depends upon a double wishbone suspension with arms 1 and 2. The linkage 4, with no fixed connection, increases wheelbase variation as a function of the wheel stroke, with benefits for comfort. In this case torque steering is not to be taken into account.

Advantages and disadvantages

Advantages

- Stabilizing toe angle variations as a function of cornering and braking forces
- Camber recovery
- Wheelbase increase at compression stroke
- Suitable for avoiding torque steering effect on rear wheel driven cars

Disadvantages

- High mechanical complexity
- Long development time
- High production cost
- High volume and weight
- High sensitivity to variation in the elastic behavior of bushings

3.3 SEMI-INDEPENDENT SUSPENSIONS

3.3.1 Twist beam suspension

Description

The twist beam suspension (Fig. 3.51) can be imagined as two trailing arms 4, fixed to the body with elastic bushing 2; the intrinsic instability of the resulting structure is corrected by a cross beam 3.

The spring shock absorber unit is fixed between arms and body.

Figure 3.52 shows further details. The arm is made from two stamped and welded steel shells. The strut will be flanged to the plate 1, welded onto the arm. The spring seat 6 is shaped on one of the stamped shells, building up the arm; the lower shock absorber mount 7 will be screwed to a tube, also welded onto the shell 5.

The cross beam 3 has a U shaped cross section and is welded to the arms at its end. The anti-roll bar is also welded to the arms.

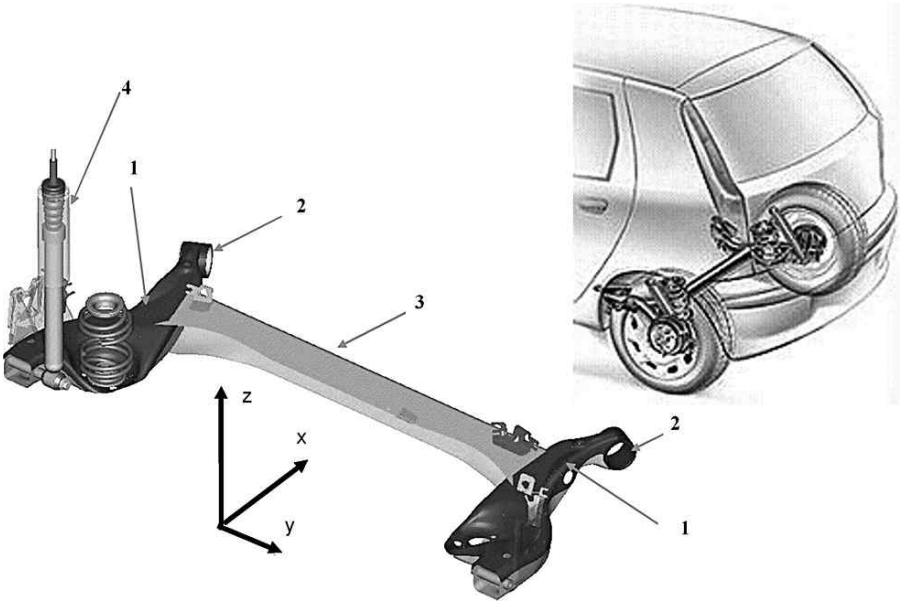


FIGURE 3.51. Installation scheme for a rear axle twist beam suspension (FIAT).

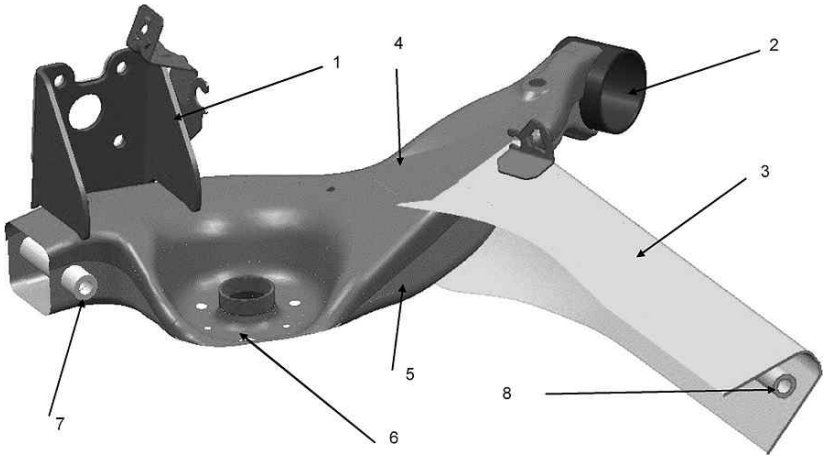


FIGURE 3.52. The longitudinal arm 4 of this twist beam suspension carries the wheel hub welded onto it. The torsion cross beam 3 has a U shaped cross section and is welded to the longitudinal arms. The anti-roll bar is welded too.

Some design alternatives for twist beam axles are shown in Fig. 3.53. Detail a refers to the so-called tubular type, where either the cross beam and the arm are made from cuts of steel, stamped and welded to each other; the twisting

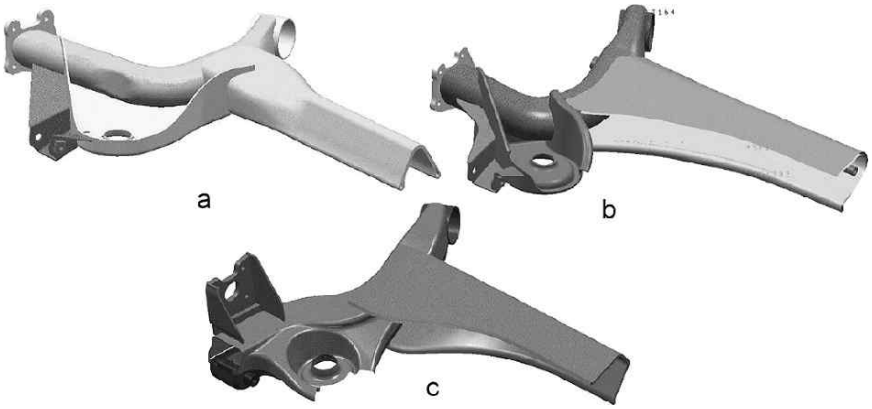


FIGURE 3.53. Design alternatives for a twist beam axle: (a) tubular type; (b) hybrid type; (c) Type made by stamped parts only.

cross beam is made by a tube crushed to the shape of a V, in order to obtain the elastic properties we will discuss later.

Detail b refers to the hybrid type, where the arm is made from a cut of tube and the twist beam from a stamped steel sheet. A possible variant of this solution can include arms made from cast iron. The choice is motivated by process optimization with regard to the existing production tooling.

By symmetric suspension stroke, arms are rotating around the axis AB (Fig. 3.54), given by joining the two center of the arms elastic bushings; there are no changes of toe and camber angles because of the suspension stroke, other than the structural deformation caused by external forces.

By asymmetric stroke, the cross beam is twisted by the difference of torque applied to the arms and, solely because of this deformation, the arms can have different angles around non-coincident axes.

To determine axes BC and AC, around which the arms rotate, it is necessary to join the bushing center of each arm with the shear center of the cross beam, in the symmetry plane of the car (point C). It should be remembered that the shape of the cross beam is such as to obtain a high bending stiffness (consider the cornering forces), with a modest, but suitable torsion stiffness. For this reason camber angle variations will be very small by symmetric strokes.

It is therefore possible to design toe angle and camber angle variations, by designing the section of the cross beam and the dimension d, relative to the wheel offset with reference to the shear center.

If the cross beam is made by a crushed tube, the V cross section of the tube can be designed to avoid the need of an anti-roll bar.

In Fig. 3.55. at left, the details of an elastic bushing are represented. The bushings axis is parallel to the car y axis; to limit toe angle variation caused by cornering forces, the inside metal body of the bushing has a tapered shape which limits this rotation.

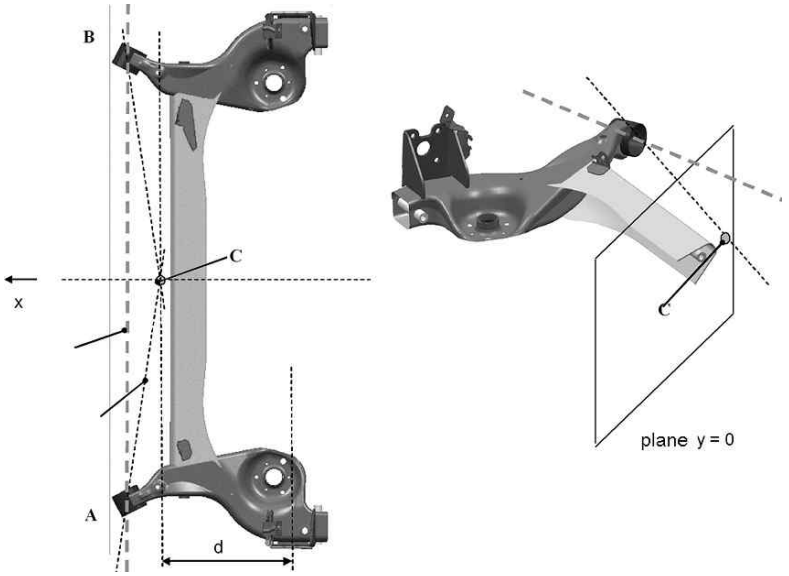


FIGURE 3.54. Scheme to determine the rotation axis of the arms of a twist beam axle in an asymmetric suspension stroke; point C is the shear center of the cross section of the cross beam.

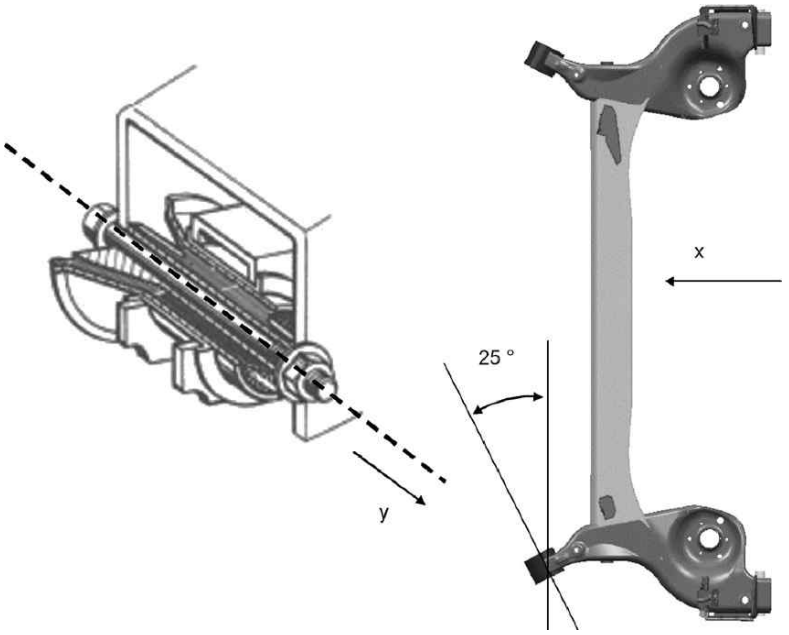


FIGURE 3.55. Twist beam axles have a natural tendency to steer because of cornering force application. This tendency can be hindered with particular rubber bushing (left) or by inclining bushing axis (right).

The same effect, suitable for obtaining understeering behavior in the car, can be produced by inclining the bushing axis on the xy plane, as shown in the right scheme.

Different arrangement are possible for coil springs and shock absorbers. The choice between the alternatives is a compromise between car comfort and the width of the cargo compartment between the wheel cases.

As a matter of fact the optimum position for shock absorbers is perpendicular to the arm as far as possible from the articulation point to the body; in this case the ratio between shock absorber and wheel stroke ratio is the highest.

On the other hand this same position brings the tip of the shock absorber inside the wheel case, reducing the cargo compartment width, particularly if the coil spring is installed coaxial to the shock absorber.

A further system to limit lateral deformation of arms because of cornering forces is to use a Panhard bar; this system is suitable for heavy cars, such as high end MPVs, when other solutions can affect comfort negatively.

Advantages and disadvantages

Advantages

- Design simplicity
- Assembly simplicity
- Reduced vertical dimension
- Almost total camber recovery by asymmetric strokes
- Possibility of controlling toe angle by body roll
- Unsprung mass smaller as with a rigid axle
- Reasonable longitudinal elasticity

Disadvantages

- High wheel case width because of expected camber variation
- Low roll stiffness
- Highly stressed parts (twist beam and its welding)
- Unsuitable for driving axles
- Toe angles too sensitive to loads
- Remarkably different behavior in empty vs. full load condition

Twist beam axles are extensively used on small and medium cars.

3.4 RIGID AXLE SUSPENSIONS

This solution, virtually abandoned in cars, finds applications in commercial and industrial vehicles and on many off road vehicles.

Applied to driving axles the suspension cross structure (axle) can integrate the final drive, differential and driving shafts.

In this case the unsprung mass is certainly heavier, as in the case of independent suspensions.

In front wheel driven cars, the rear idle axles have a simple structure connecting the two wheels; the unsprung mass value is not far from the range of independent suspensions.

Looked at kinematically point of view, this solution features no track variation by roll and parallel springing motion. Wheel camber referred to the ground does not change because of body roll.

We will consider two different design alternatives: Rigid axles with leaf springs, where the kinematic function is also performed by springs, and the guided rigid axle, where the kinematic function is performed by dedicated linkages, as with independent suspensions. In this case the elastic elements can vary as with leaf springs.

3.4.1 Rigid axles with leaf springs

Figure 3.56 shows the drawing of a simple solution suitable for front wheel driven vehicles; in this case it is a small commercial vehicle.

The two wheels are connected by the axle 1, whose shape is created by the bulk of the other under body components, in this case, the spare wheel.

The axle is linked to two leaf springs with flanges and screws. The spring is connected to the body so as to be able to change its length, because of the suspension stroke; the front end is connected to a fixed eye 3, where the rear end is connected through two *swinging shackles*, a kind of small pendulum rod. The stop spring 5 works on the axle directly, as shock absorbers 6.

The leaf spring is bent on the zx plane; curvature is positive when the center of curvature is above the axle. This curvature is peculiar to axles that are self-steering by body roll, and is also necessary for obtaining the desired value of suspension stroke.

If curvature is positive the steering angle of the axle oversteers the car; if negative it understeers and gives. Flat springs are neutral. Because curvature decreases as load increases, this contribution becomes more understeering as the load increases.

This effect has to be added to that of leaf spring lateral deformation, which is normally oversteering.

In our example the leaf spring shows a single leaf, with advantages for suspension hysteresis.

This simplification is not always possible, however, because of the weight of the vehicle.

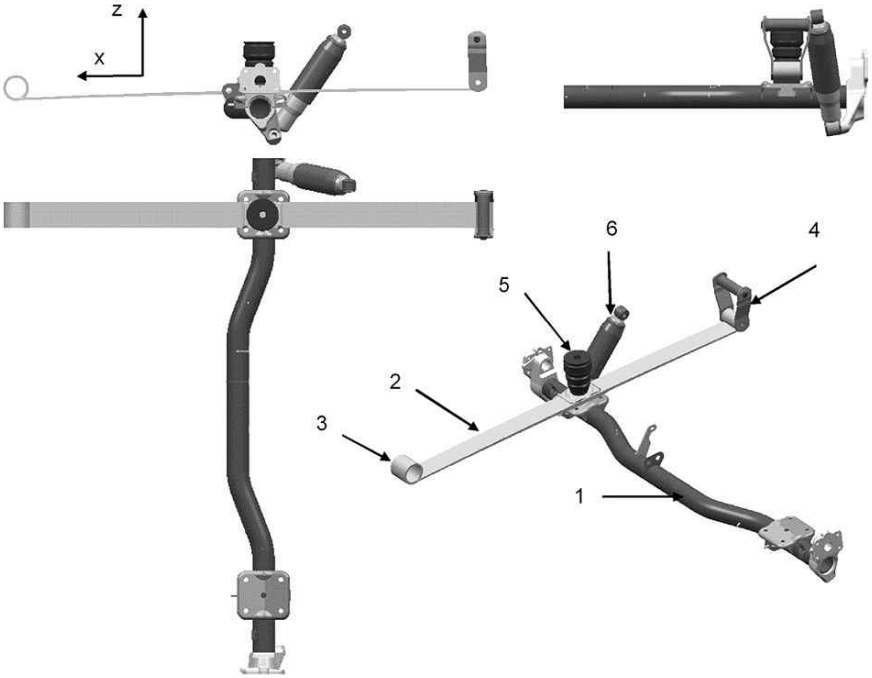


FIGURE 3.56. Simple rigid axle suspension with single leaf springs for a small commercial vehicle (FIAT).

When the leaves are numerous they feature different lengths; they should be free to slide over each other, in such a way as to build a body of uniform flexural resistance. The stack of leaves is, in fact, with semi-elliptic leaf springs, bound to have the same curvature and the resulting structure can be considered as a different arrangement of a beam with a rhomboid shape, supported at the ends and loaded in the middle, as shown in the lower part of Fig. 3.57.

In some cases single leaf springs are used, as in our example or in Fig. 3.57 at right; in this case the leaf is tapered according to a theoretical parabolic shape, to again obtain a body of uniform flexural strength. In other cases tapered leaves can be joined in twin assemblies as in the upper right. If the fabrication process is suitable to produce leaves with variable thickness, stacks can contact at the leaf ends only, with benefits for hysteresis.

The central part of the stack is fixed to the axle through U shaped brackets (see Fig. 3.61 in the following paragraph); other brackets keep the leaves in position.

The longest leaf has two eyes, sometimes copied by the second leaf, for attachment to the body. In semi-elliptical leaf springs, layers of self-lubricating plastic strips are interposed between the leaves. These decrease friction and avoid leaf sticking because of corrosion after long stops.

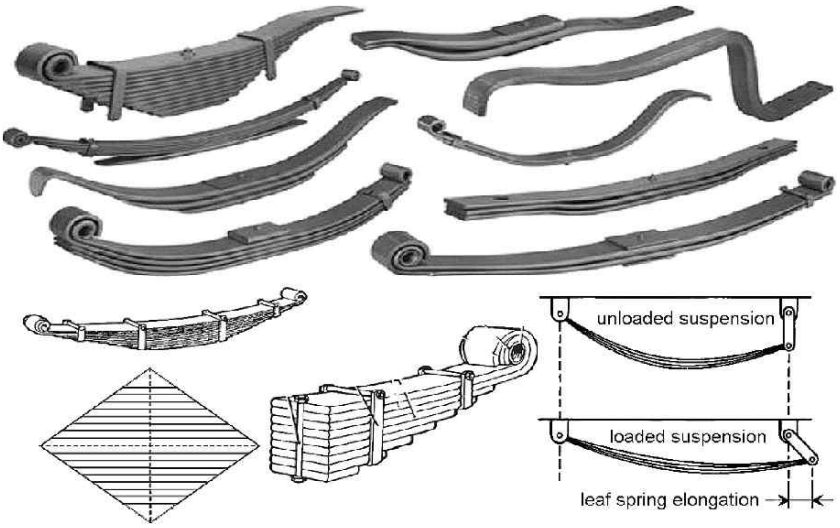


FIGURE 3.57. Different kinds of leaf springs for rigid axles (upper part); the scheme shown on the left at bottom shows how a semi-elliptic leaf spring can be seen as a uniform flexural strength body; the scheme on the left shows the motion of the leaf spring end because of suspension stroke.

Advantages and disadvantages

Advantages

- Extreme design simplicity.
- The only suspension with full camber recovery by roll.
- Robustness suitable to off road and heavy duty applications.
- Wheel stroke is greater than spring stoke by roll motion; this property is useful in off-road vehicles to clear transversal trenches.

Disadvantages

- High elevation of the body floor on the cargo compartment.
- High restriction on wheel cases reducing cargo compartment width.
- Heavy unsprung mass.
- Wheels working angles cannot be adjusted.
- Poor elasto-kinematic characteristics.
- Low longitudinal elasticity.
- Low roll stiffness.
- Vibrations caused by S deformations of the leaves.

3.4.2 Rigid guided axles

To improve the transversal and longitudinal guidance of the axle and to obtain better elasto-kinematic behavior, the axle can be made more complicated by adding suitable linkages.

These additional linkages are essential if the springs, such as coil springs and pneumatic springs, are only capable of absorbing vertical loads .

Figure 3.58 (scheme A) shows a solution where the reaction to cornering forces is assigned to a Panhard bar, linked to the car body at one end (point 2) and to the axle at the other (point 1). If coil springs are adopted other linkages, such as longitudinal parallelograms, are necessary, not visible in this figure, to build up the necessary longitudinal guidance.

Panhard bars can also reduce the natural oversteering behavior of a leaf spring rigid axle, setting limits to lateral spring elasticity.

Scheme B shows a solution where the vertical guidance of the axle is managed by a Watt quadrangle. Here the axle is usually shaped as a rigid triangle, where the basis connects the wheel hubs and the vertex is set on the symmetry plane of the body in front of the wheels, connected through a spherical joint. This scheme is named the De Dion axle.

The Watt quadrangle, by its shape, obliges the middle of the axle to move in the symmetry plane of the body. The front spherical joint can receive a certain flexibility so as to attain a given value of understeering.

Figure 3.59 shows possible linkages for a rigid axle in connection with the application of coil springs.

Scheme A shows two longitudinal arms and an anti-roll. Arms react to longitudinal loads, while their inclination can also absorb cornering forces; the anti-roll

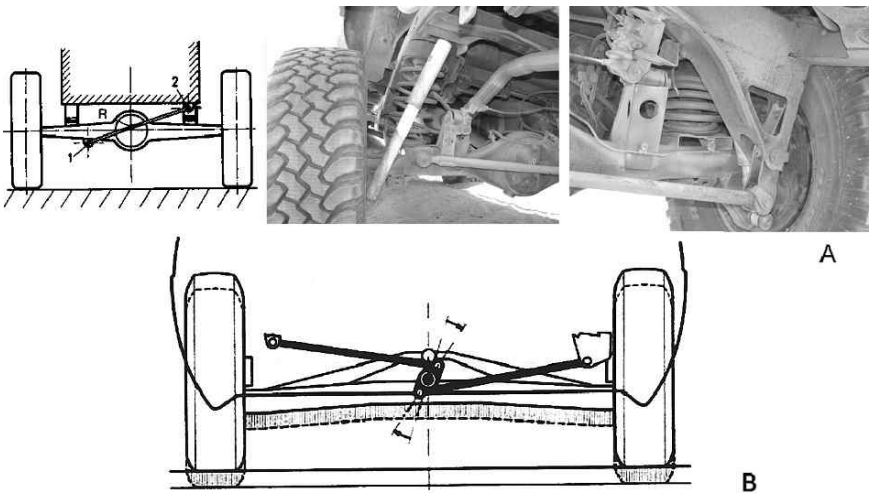


FIGURE 3.58. Systems for axle guidance; at top a Panhard bar; at bottom, a De Dion mechanism.

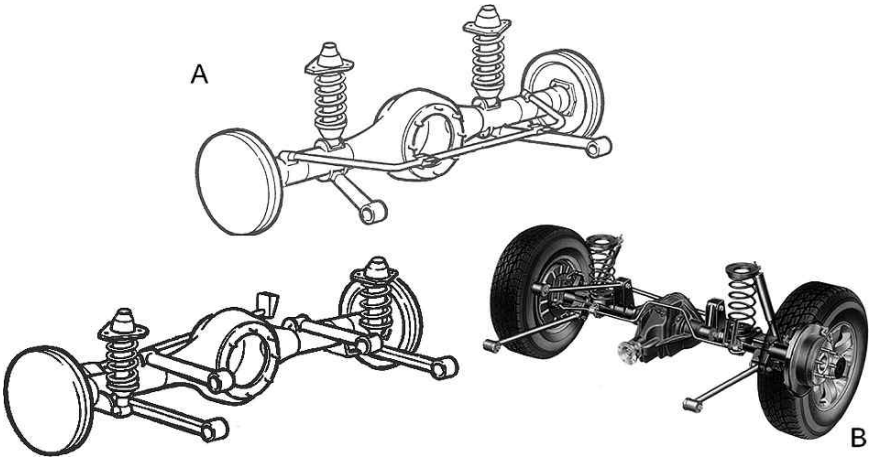


FIGURE 3.59. Different rigid axle linkages suitable for coil springs; the anti-roll bar itself (upper scheme) can work as a functioning linkage.

bar can take longitudinal loads, contributing by reacting to the torques applied to the axle.

Scheme B provides four longitudinal arms and a Panhard bar. The longitudinal bars react to longitudinal forces and driving and braking torques, while cornering forces are absorbed by the Panhard bar. In a possible variant to this version use of the Panhard bar can be avoided by a suitable inclination of the longitudinal beams. In both cases the rigidity of the elastic bushings can be designed with an assigned elasto-kinematic behavior.

3.5 INDUSTRIAL VEHICLES SUSPENSIONS

Industrial vehicle suspensions must allow notable load variations particularly on the rear axles; bear in mind the ratio between curb weight and full load on a truck. This ratio is lower by an order of magnitude than that of cars.

Industrial vehicles have a structure including a chassis frame made by side beams and cross beams, bearing chassis components, cabin and pay load.

Front and rear suspensions on trucks are usually of the rigid axle type, similar to those introduced in previous paragraphs. Buses and commercial vehicles also adopt independent suspensions.

Many vehicles of this kind adopt pneumatic springs instead of leaf springs. The advantage of this choice is the possibility of controlling vehicle trim and the springing characteristic as a function of the actual load.

Their application implies an increase of cost and a more complicated axle mechanism, because the guidance function of the leaf spring is missing; the necessity of compressor and storage pressure bottles is partially overcome because they are already demanded by the braking system.

In light industrial vehicles, are still widely found leaf springs.

Sometimes in order to increase the suspension stiffness at higher loads progressively, more leaf springs set in parallel are applied, as shown by the example on the second left row of Fig. 3.56. The spring set in parallel works when the payload exceeds certain limits; this reduces vehicle trim variations and approximates the constant natural frequency condition more effectively.

3.5.1 Pneumatic springs

Air springs allow control vehicle trim to be changed and controlled. They also reduce the load excursion effect on suspension geometry, with benefits for comfort and, on trucks, for the integrity of transported goods.

As will be explained later, minimizing sprung mass acceleration implies a spring stiffness proportional to the sprung mass (constant natural frequency).

Because of the degree of payload variations in industrial vehicles, spring rigidity should be highly progressive. But springs are designed for maximum payload and prove to be too stiff at partial payload or in empty vehicles.

Air springs overcome this limitation, by changing their inflation pressure.

Figure 3.60 shows some examples of pneumatic springs and a simplified scheme of their installation on the vehicle. The spring is a rubber bellow reenforced with textile fibers, like a tire, containing pressurized air.

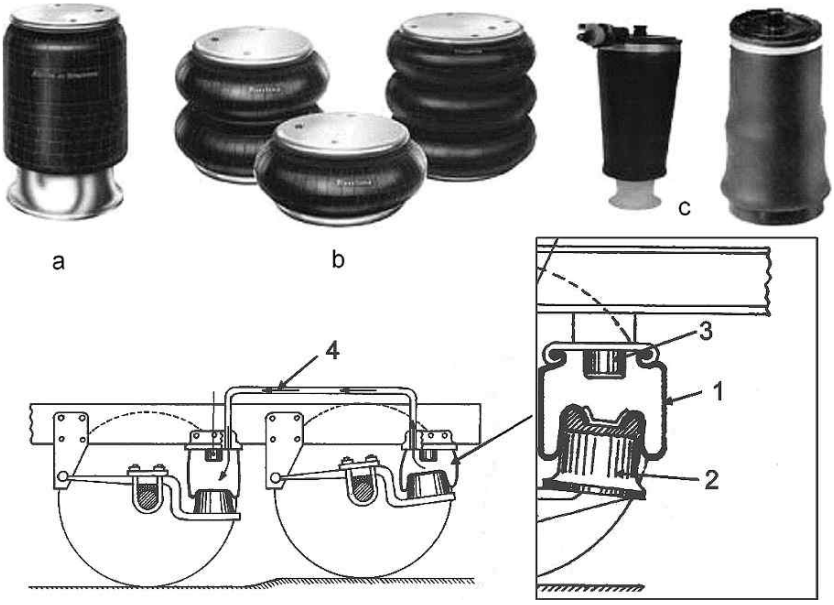


FIGURE 3.60. Different kinds of pneumatic springs for the suspension main spring; if the elastic elements of tandem axles are pneumatically connected, the two axles build up an isostatic system easily (lower scheme at left); stop springs are usually included inside the rubber element (lower scheme at right).

Types a and c show a bellow shaped like a tube 1 (lower detail), which can roll up on a metal surface 2 to adapt for different lengths; this configuration is suitable for light vehicles where the spring diameter is reduced with reference to the stroke; the bellow diameter, at the same inflation pressure, must be proportional to the square root of the sprung mass.

For type b the bellow is reinforced with metallic rings; this shape is suitable for heavier vehicles where the diameter has the same magnitude as the useful stroke (notice that this picture is not scaled to the others).

Air, with its characteristic of changing the volume as a function of pressure, is the elastic element of the suspension. With a compressor and controllable valves it is possible to adjust the pressure to keep the same vehicle trim at any load.

Pressure variations can also adjust vehicle height when needed (for instance to lower the saddle of a tractor when hauling a semi-trailer).

The physical transformation of the air can be assimilated to an isotherm during slow trim variations; it must be assimilated to a polytropic or adiabatic transformation during springing motion, where the time allowed for thermal exchange is low.

In this last case, the spring characteristic provides no constant flexibility, possible effective diameter changes notwithstanding.

The detail also shows how the rubber stop spring can be integrated into the pressure bellow.

The lower scheme shows how it is possible to obtain an interconnected suspension on a tandem axle. Tandem axles are used to limit road loads in heavy vehicles. The interconnection eliminates the intrinsic hyperstaticity of the system.

Similar springing elements can also be used in light vehicles and cars when the vehicle trim must be adjustable, such as in off road vehicles, for instance, where the ground clearance is increased for off road driving.

In a dedicated chapter we will also discuss the control techniques applied to these systems.

3.5.2 Front suspension

Rigid axles

In European industrial vehicles the steering wheel is usually set over the front axle or in the front overhang.

This kind of architecture is, therefore, characterized by the application of a longitudinal steering rod connecting the steering box with the steering knuckle, as can be seen in Fig. 3.61.

As the figure shows, the steering box of rack and sector type with ball recirculation rotates the drop arm 1, articulated to the longitudinal steering rod 2. This last is connected to the steering arm 3, part of the steering knuckle,

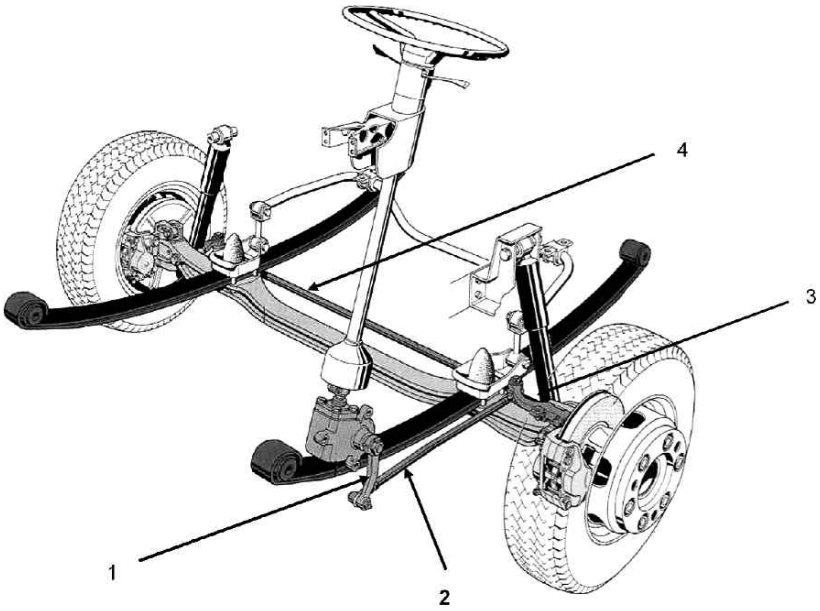


FIGURE 3.61. Steering mechanism for a front rigid axle of an industrial vehicle (Iveco).

or stub axle. A cross steering bar 4 connects the right steering knuckle, so as to obtain correct steering.

To limit the steering angle consequent to the suspension stroke, the articulation between the drop arm and longitudinal steering rod must be set at the instantaneous rotation center of the axle, or leaf spring, relative to the body, close to the front eye of the spring.

Figure 3.62 shows a rigid steering axle with its main components. The leaf spring features two equal length tapered leaves fixed to the chassis frame in the established way. Leafs contact each other at the ends and in the middle only; in this way the behavior of a uniform stress body is obtained, with limited friction and hysteresis.

A third very short leaf can increase the flexural stiffness of the spring assembly for high loads and decrease S deformation during braking.

The S deformation is called in this way because of the shape of elastic deformation of the leaf spring, created by the application to the axle of a pure torque.

Such deformation is particularly annoying because the moment of inertia of the axle around the y axis, can lead to periodic deformations, where the axle accepts a torque higher than the friction limit for a short time and then slips when the inertial force disappears. The result is vibration and loss of braking efficiency or traction.

Leaves are fixed to the axle 3 through U bolts. The anti-roll bar 6 moves with the axle (and therefore must be accounted as unsprung mass), while its

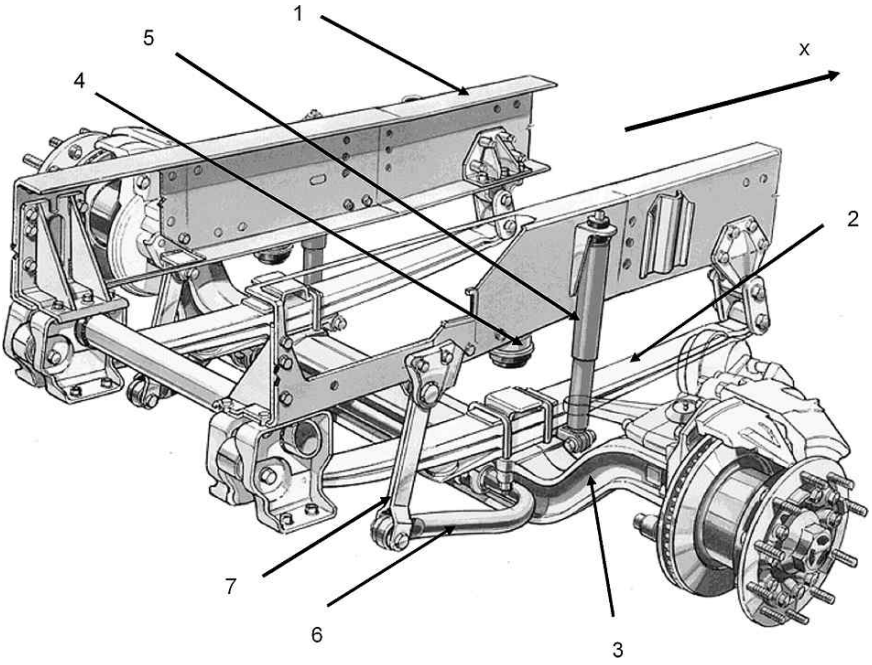


FIGURE 3.62. Design details of a front rigid axle suspension for an industrial vehicle (Iveco).

ends are connected to the chassis frame through pendulum bars 7; rubber stop springs 4 work directly on the axle.

The drawing in Fig. 3.63 shows a cross section through the front wheel axes. The king-pin axis and king-pin offset are identified by the rotation axis of the steering knuckle. The wheel disc has a bell shape, necessary to contain the king-pin offset considering the space necessary for the large disc brake; this shape is standardized for all vehicle wheels and must take into account the fact that by rear twin wheels tires should not contact each other or trap rubble.

As we explained, this axle in cars has the advantage of keeping the wheels always perpendicular to the ground; on such large tires camber angles are not acceptable.

A further advantage is the possibility of placing each suspension component below the chassis frame, simplifying its shape and adding space for the payload or cabin.

A disadvantage is that it is difficult to control the elasto-kinematic characteristic of the suspension because of the rigid connection between the wheels.

On bumpy roads the wheels can be steered by the suspension stroke. Figure 3.64 shows the path of the leaf spring connection point to the axle, during suspension motion; this point moves forwards during suspension extension and back during suspension compression. By parallel springing the axle should

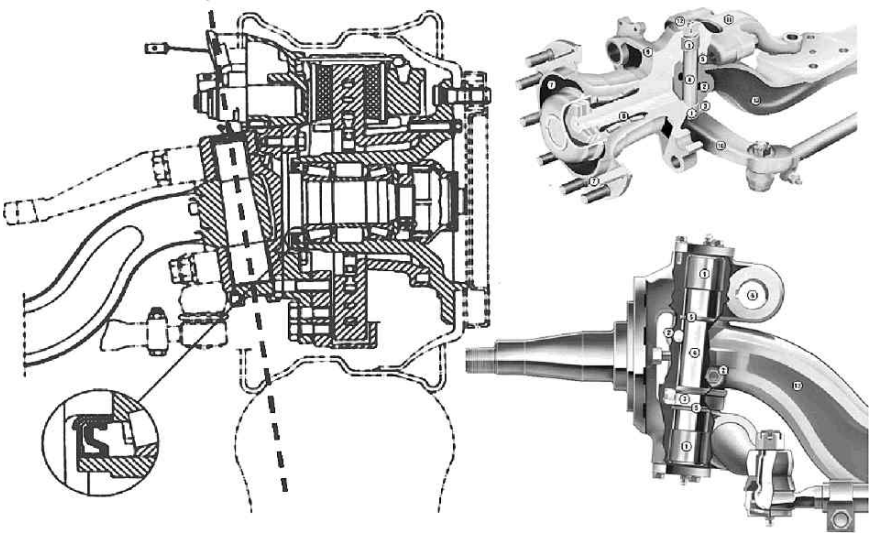


FIGURE 3.63. Stub axle of a rigid steering axle, showing the king-pin axis position.

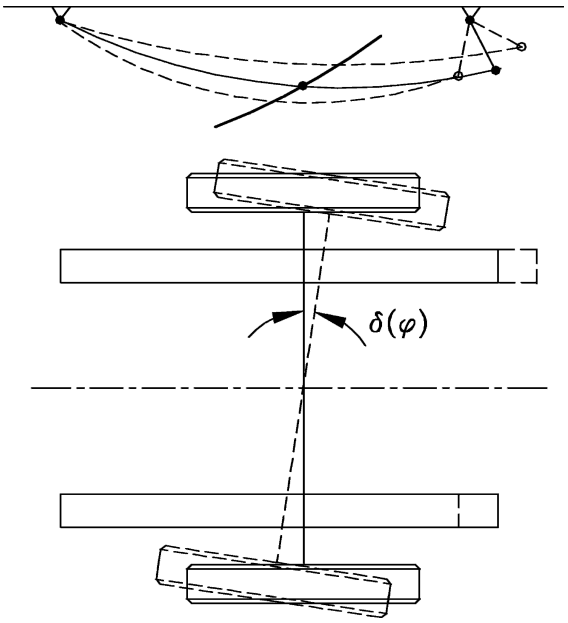


FIGURE 3.64. Roll steering of a rigid axle.

not steer if the longitudinal steering rod articulation is set in the centre of curvature of this path. A certain steering angle will be created by asymmetric springing (body roll).

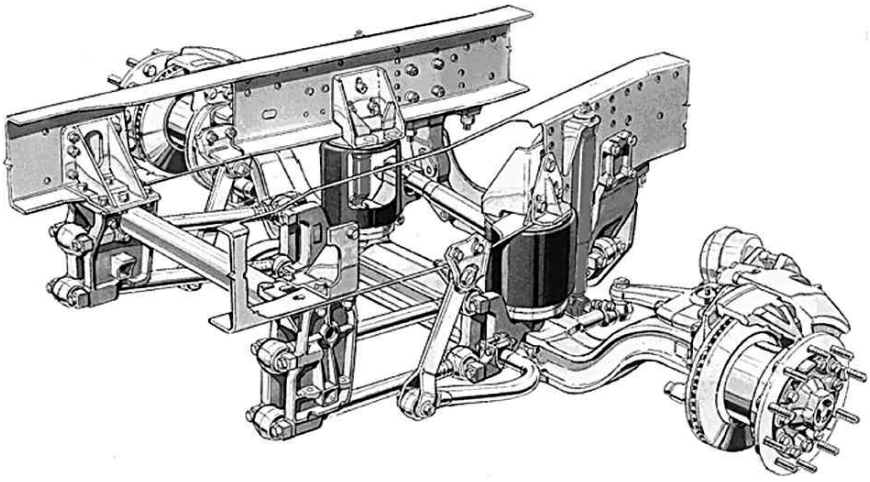


FIGURE 3.65. Rigid axle suspension for an industrial vehicle; the application of pneumatic elastic elements is made in conjunction with longitudinal and transversal guiding linkages (Iveco).

Guided rigid axles

Figure 3.65 shows a rigid axle suspension where the axle is guided by two longitudinal parallelograms, similar to those we have seen for small vehicles.

There are two pairs of longitudinal rods dedicated to reacting to longitudinal forces and torques. A Panhard bar is dedicated, as usual, to cornering forces. Springing is obtained through pneumatic bellows.

The advantage of this arrangement is better control of the axle steering in roll and the possibility of applying pneumatic springs. This kind of mechanism can also be applied with leaf springs, which will be free to move longitudinally at both ends; this mechanism avoids the S deformation completely.

Double wishbone suspensions

The need for better ride comfort for bus passengers and better handling performance due to increased speed has led industrial vehicle designers to apply to these vehicles a front wheel double wishbone suspension, similar to those applied to cars.

A further need for these industrial vehicles is to keep the passenger compartment floor as low as possible in order to make boarding easier; this is particularly important for urban buses, where there is a frequent exchange of passengers. A rigid axle causes too high a position for the chassis frame making the double wishbone suspension advantageous.

Figure 3.66 shows two examples of this suspension. Air springs are applied on both suspensions; the suspension mechanism scheme is the same used in cars.

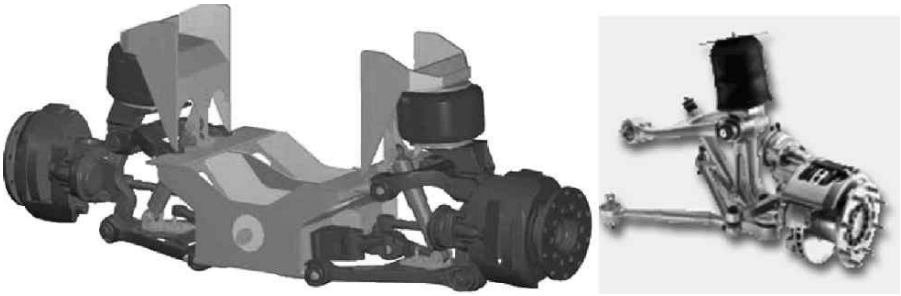


FIGURE 3.66. Front double wishbone suspensions with pneumatic spring for buses.

Air springs allow urban buses to incline the body by stops, in order to lower the entrance door further.

3.5.3 Rear suspensions

The need for twin tires to be always perpendicular to the ground precludes suspension configurations other than rigid axles.

The high value of the load spread on heavy duty vehicles has practically standardized the application of pneumatic springs; in tractors the vehicle can be lowered when the semi-trailer must be attached.

Figure 3.67 shows a rigid axle suspension architecture widely applied to heavy duty industrial vehicles. The axle is guided by two lower longitudinal bars 2, which react to longitudinal loads; the upper triangular arm is connected to the axle through a spherical joint 1 and reacts to cornering forces and torques.

Pneumatic spring bellows 4 are double for each side to allow standardization with those in front.

On highly loaded vehicles tandem axles must sometimes be applied in order to comply with legislative load limits on a single axle.

Tandem, or multiple axles must distribute the load to the ground in a uniform way and adapt to uneven roads or bumps.

An isostatic condition is also necessary to better distribute the reaction to the payload when this is not uniformly distributed.

An isostatic condition is easy to be satisfied through air springs by a suitable interconnection of pressure air pipes.

Tandem axles using leaf springs obtain this condition using two cantilever springs free to rotate in their middle section, as Fig. 3.68 shows.

The axles are suspended by symmetric longitudinal triangles and bars, as in the previous figure; this mechanism causes the two axles to carry the same vertical load.

Leaf springs cannot react to pitch; front shock absorbers have to limit by a suitable stop rubber spring any excessive leaf spring rotation caused by driving on bumps.

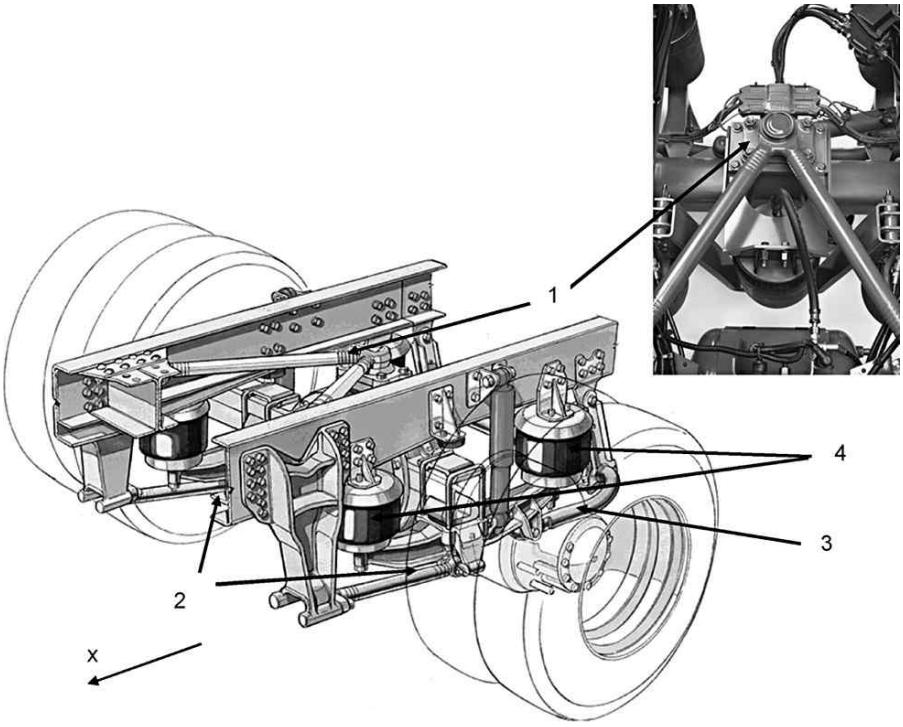


FIGURE 3.67. Rear suspension for a heavy duty industrial vehicle, featuring pneumatic springs; the axle is guided by an upper triangular arm and two lower longitudinal arms (Iveco).

3.6 DESIGN AND TESTING

3.6.1 Design preliminary outline

The design preliminary outline of a suspension scheme consists essentially of the determination of the geometric coordinates of the articulation points and of the position of the main axes, such as king-pin axis, shock absorber axis, etc. The goal is to reach the assigned target for elasto-kinematic performance.

The choice of suspension scheme has been made previously, taking into account the volumes of the mechanical components of the chassis, the minimum clearance from the ground, the body structure and the other conditions determined by costs and available production technology.

Suspension architecture is also one of the selling points for automobiles.

In this paragraph a step by step logic flow of the design outline operations is explained for a McPherson suspension; the extrapolation to other suspensions is left to the reader.

We refer to the case of a front steering axle by a transversal power train.

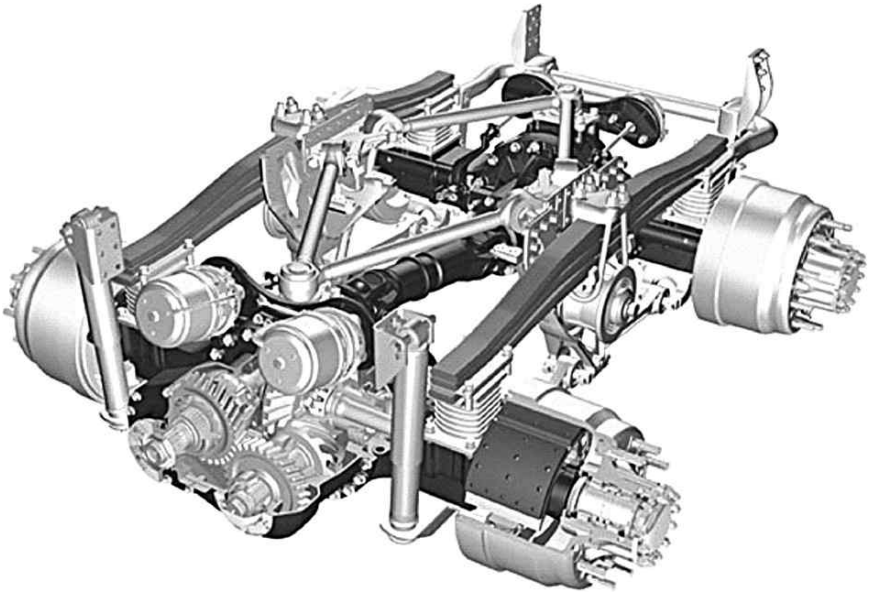


FIGURE 3.68. Cantilever leaf springs allow an isostatic tandem suspension to be obtained. Leaf springs are connected to the frame by a pin in their middle and axles are guided by longitudinal bars and triangles (Iveco).

The first step consist of the identification of the king-pin axis position. This should be consistent with what has been decided for king-pin offset and trail in relation to steering wheel feel and return and braking stability.

In a McPherson suspension the king-pin axis is determined by the center of the spherical joint A and the center of the upper pivot B (Fig. 3.69). Point A is set along the king-pin axis in such a way that the spherical joint could find its place between the constant velocity joint and the wheel rim.

To reduce as much as possible the reaction to lateral forces on the lower arm, the point A is set as low as possible. The lower limit is set by wheel rim and braking disc positions.

The point B position must allow the assigned suspension stroke and correct operation of the coil spring and stop spring.

The elevation of point B must be as low as possible to comply with a low hood profile; hood profile has an influence on style and aerodynamic performance. The clearance between hood profile and suspension upper bearing depends upon requirements of passive safety related to impacts against pedestrians.

When the king-pin axle is set, the shock absorber axis can be set. This joins points B and C, on the piston rod of the shock absorber, as shown in Fig. 3.70.

To avoid mechanical complication of the strut design, point C, as seen from the y direction, should be set on the king-pin axis.

The position of point D, setting the rotation axis of the lower arm, is a consequence of the desired camber recovery and track variation, as a function of

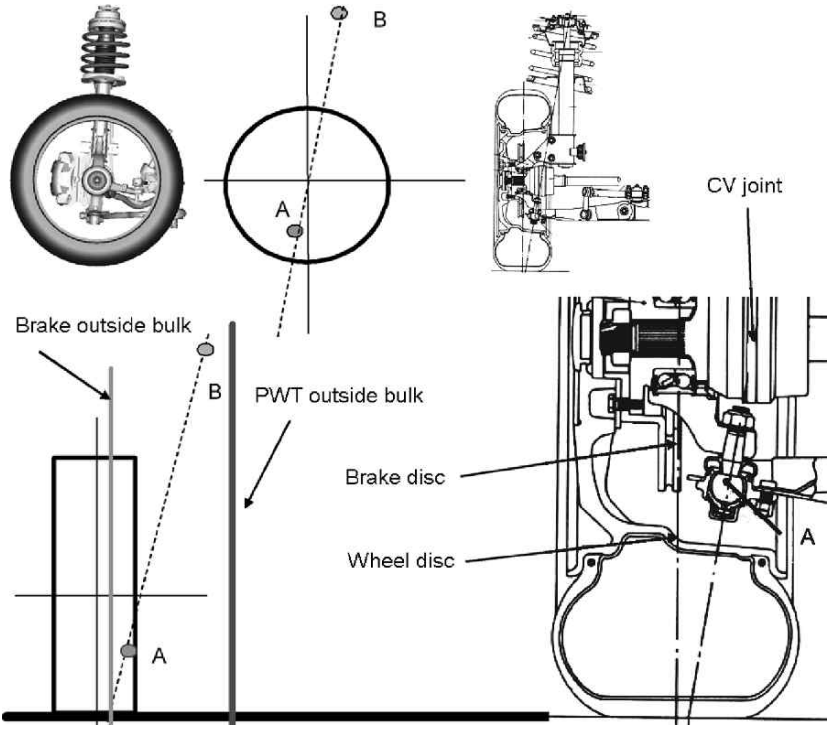


FIGURE 3.69. Points A and B positions of the king-pin axis are conditioned by power-train bulk, on one side, and brake disc bulk, on the other side, and by constant velocity joint position.

suspension stroke (see Fig. 3.70). D is the intersection of the arm rotation axis with a zy plane determined by the point A position.

Because of the kinematics of this mechanism, to increase camber recovery, point B must be closer to the inside of the car or the shock absorber inclination must be increased; these positions must be set with attention to possible mechanical interference and the position of the king-pin axis, which limit camber recovery in practice.

As an alternative, camber recovery can be improved by raising point D. This modification implies, as we will see later, the elevation of the suspension roll center, with impact upon other vehicle performance (Fig. 3.71).

Now the rotation axis of the arm can be determined by setting points E and F, the centers of the elastic bushings.

Most solutions provide the same y coordinate for these two points; we will adopt this assumption for sake of simplicity.

The z coordinates of these points are chosen to limit pitch angles during acceleration or braking; we will explain later about the anti-dive properties of their positioning.

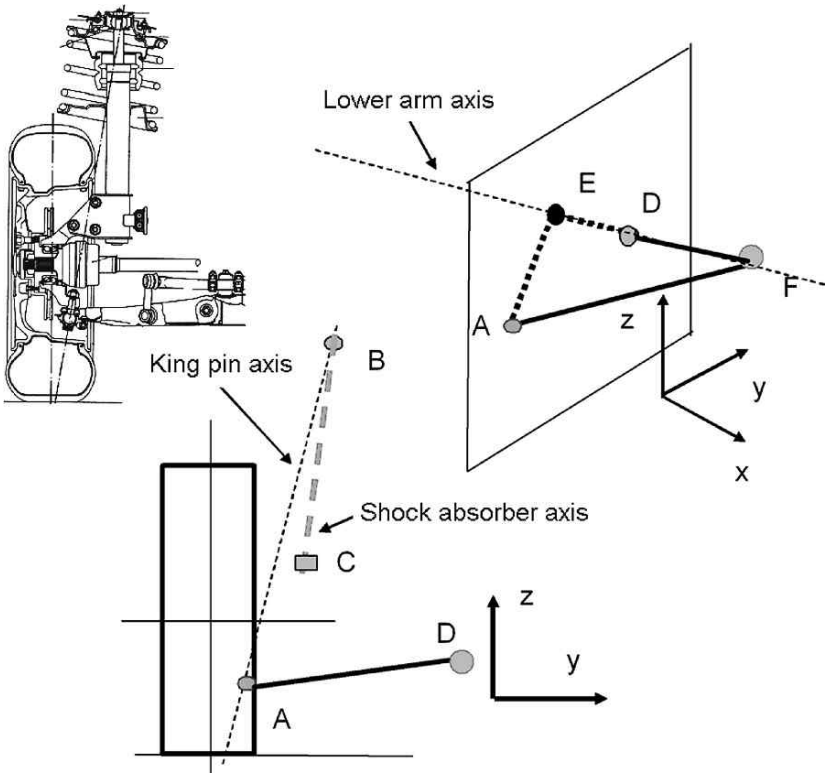


FIGURE 3.70. The arm rotation axis EF is determined to obtain the desired camber recovery, once A and B are determined.

The x coordinates of these points determine the shape of the arm and are determined by the space available for the articulation points and the kind of subframe that will be adopted.

Their position also has direct consequences on the value of the forces that are applied to points E and F, by longitudinal shocks against obstacles and by cornering forces.

By combining in a suitable way the position of these points and the radial and axial stiffness of the elastic bushings, it is possible to control longitudinal wheel flexibility, impacting comfort, and camber angle variations by cornering forces, impacting tire cornering stiffness and, therefore, handling.

At the end, the positions of the articulation of the steering mechanism will be set, as shown in Fig. 3.72; additional explanations on this subject will be given in the chapter dedicated to the steering system.

The choice of point H position (spheric joint between strut and steering rod) is determined considering the Ackermann correct steering condition, the torque permissible on the steering wheel and the position available for the steering rack.

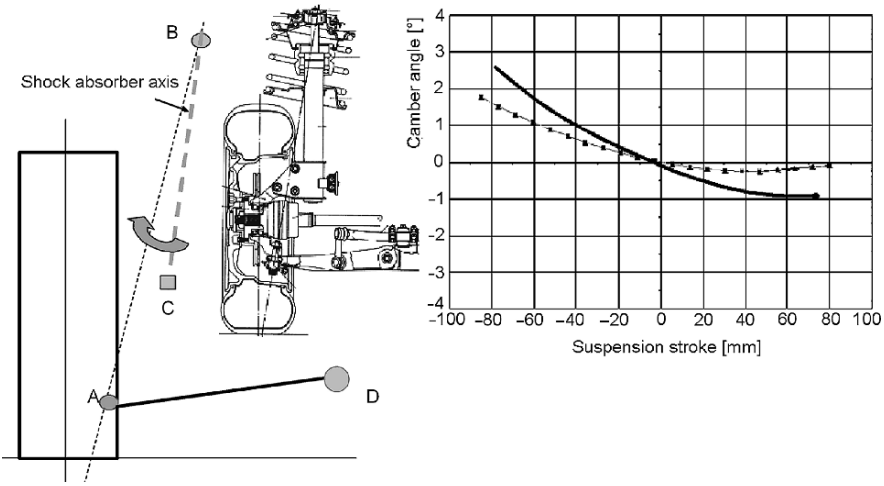


FIGURE 3.71. A correction in the position of point D can change the shape of the camber angle variation diagram as a function of the suspension stroke.

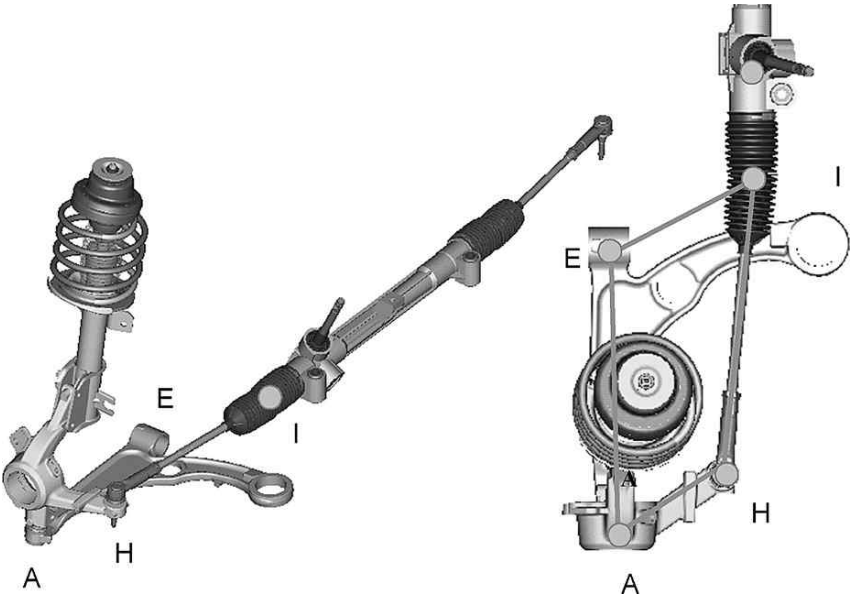


FIGURE 3.72. The position of point E should allow the wheel to move back against an obstacle, without noticeable variations of the toe angle and camber angle.

Point I (spheric joint between steering rod and rack) is determined in order to limit toe angle variations because of the suspension stroke. In a view from the z axis, points A, E, I, H must shape a parallelogram in order to limit toe angle variation caused by longitudinal motions of the wheel.

After the preliminary design outline of the suspension, the detailed design of suspension components can begin, taking into account the structural purpose of this part of the chassis. This undertaking can be considered from two different points of view:

- Mechanical integrity, defined by the fatigue and shock resistance of the suspension components
- Deformations under the applied loads or elasto-kinematic behavior

We are going briefly to analyze these two approaches.

3.6.2 Structural integrity

Fatigue behavior

Fatigue behavior may be evaluated, through calculations or tests; in this latter case dedicated facilities are used, such as benches and measurement systems. In both cases, reference is made to load specifications that simulate in a synthetic way the entire operational life of the component under investigation.

The load synthesis applies scientific methods to real load histories on road circuits chosen according to the manufacturer's experience and the purpose of the vehicle.

Structural analysis can be performed on the complete suspension or on single components. In the first case it is good practice to include in the test that part of the body or subframe structure where the suspension is connected.

The suspension under test is usually installed on a bench including a real part of the car body; this dummy proves to be the less expensive installation fixture to be used. The wheel hub is linked to a three axis actuator, applying loads in the three directions of the car reference system.

Load and displacements, which are applied, are extracted from on road measurements applicable to the average life of the vehicle; at regular intervals of time the general integrity of the suspension is checked (cracks, bushings or bearings deterioration, etc.).

The mission load definition is made, as we have said, by recording forces F_x, F_y and F_z and moments M_x, M_y and M_z applied to the wheel hub on an assigned road circuit, where a similar vehicle equipped with measurement and recording devices is driven.

It is obviously important do define the minimum life, in term of distance travelled, necessary on a given course; different courses can be travelled in sequence.

When the suspension under test has reached its target life, the following evaluations are made:

- Bolts and screws thrust decay

- Main characteristics of disassembled components, to be compared with those at the beginning of the test, with particular reference to bushings, bearings, springs, stop springs, spherical joints and shock absorbers

Usually the suspension is assembled again and the test is continued until at least three failures are reached or the life is doubled.

The suspension target life must be at least the same as that of the entire vehicle, usually 200,000 km for a car and in the range $300,000 \div 800,000$ km for industrial vehicles, according to the kind of application.

Broken components or those not accomplishing their design specifications are analyzed and modified until they reach the desired life.

When single components are tested, forces and moments are applied to their mounting interfaces. These are calculated with mathematical models beginning with forces and moments applied to the wheels. This test procedure is particularly cost effective when only single components are modified on a suspension or when tests must be repeated to gather a sufficient confidence level.

In this case simplified tests are defined, which are normally more severe and which the experience has demonstrated to be equivalent to the baseline test.

As an example of this approach, we can examine a suspension arm; by tests or calculations the component is linked to a testing block and loaded with wheel forces.

The test cycle can include:

- Longitudinal loads, due to braking forces (at least 3,000 load blocks representing a stop brake with the highest friction coefficient)

- Cornering forces, due to curves (at least 300,000 load blocks, each simulating a bend with the highest friction coefficient)

This example does not factor in loads due to road irregularities, because they are not as relevant to the suspension arm.

If the suspension is applied to a driving axle, traction load at maximum friction coefficient must also be taken into account.

If the part under test is affected by vertical loads, maximum loads applied by stop springs are also considered; usually they are assumed to correspond to a vertical acceleration of 35 $\mathrm{m/s^2}$.

For a more accurate evaluation of component life stresses of the most highly loaded parts are calculated with analysis codes, their results interpreted using a fatigue damage criterion.

The scheme of this method is the following. Let us consider that this component undergoes n different load histories ($P_k(t)$, with $k = 1, n$). P_k is a generic load applied to a point of our structure.

Using the stress analysis code stresses are first calculated, generated by the forces applied to the same points ($\sigma_{ij,k}(x)$, with $k = 1, n$) where loads P_k are applied.

The fatigue code can calculate the time history of stresses by a linear superimpositions of the effects, as:

$$\sigma_{ij}(x,t) = \sum \sigma_{ij,k}(x) P_k(t).$$

Suitable algorithms allow us to summarize any three axial stress situation in a single axis under the same stress; for metal components the distortion energy criterion is usually applied.

Afterwards the fatigue life of each point of the structure can be evaluated, with reference to the material characteristics.

The Miner hypothesis, stating that a body that undergoes different stress levels reaches the end of its life (rupture) when the sum of the partial damages of each of these levels reaches the value of 1, is frequently used. The partial damage is, by definition, the ratio between the number of cycles actually performed and the number of cycles necessary to reach a rupture at a given stress.

Defects of the material and shape oddities must also be taken into account.

A simple way to evaluate these effects is to evaluate material fatigue by testing samples extracted from parts of the same components or similar components produced by the same process.

In this way many Wöhler curves can be achieved, suitable for evaluating the influence of the main factors on the material fatigue life.

Misuse

It is common practice to take into account a reasonable *misuse*: Misuse is an incorrect, but possible use of the component. The idea here is that, after one of these misuses, the driver understands he has made a mistake and has learned never to repeat it, but it would be unreasonable to have the mission interrupted solely because of this mistake.

A minimum number of misuses must be allowed without rupture; in the case of suspensions misuse usually implies the application of shocks.

Certain obstacles must be encountered without damages or permanent deformations; a typical test example consists of driving at 50 km/h over an obstacle 100 mm high and 100 mm long.

A different kind of misuse test is hitting a wheel against a sidewalk, as could happen when driving a curve at an excessive speed.

In this case the structural integrity of the suspension cannot be guaranteed, particularly for the lower arm; occupant safety must be, instead, protected within certain limits of acceptable severity.

After a maneuver of this kind, the following results must be obtained:

- Arms must not brake so as to maintain their guidance to the wheel until the car is stopped.
- Permanent deformation on the arm must be visible and such as to discourage any continuation of the trip.

Crash test

In automotive engineering crashes against external obstacles or other vehicles make reference to a standard test specified by law.

Suspensions must be designed as to not intrude into the passenger compartment as a result of a standard crash test.

These structures must prove to collapse in a non-aggressive way; at the same time subframes, wheels and other elements have to play an important role in the entire car's structural collapse, after a crash test, to conserve the geometrical integrity of the passenger compartment.

3.6.3 Elasto-kinematic behavior

After the explanation of different kinds of suspension and the analysis of their advantages and disadvantages, it is clear that a certain technical hierarchy could be established, ranking suspension types by increasing function but also increasing cost.

Front suspensions could be ranked starting from the McPherson to double wishbones with virtual centers, while rear suspensions could be ranked from trailing arms and twist axles to multilink.

If the designer decision is easy for a high performance car with no restrictions on price, or for a small economy car, where costumers are keen to sacrifice for a lower price, the choice is more difficult for most medium cars.

Good results can often be achieved by a cheaper suspension if the design is sufficiently refined, particularly as far as tire sizing and elastic bushings are concerned.

Design rules

Some design rules can be set to address suspension design to obtain good handling and comfort.

Tires working angles

Let us begin with tire working angles and their impact on handling.

If it is true that working angle variations, in consequence of body roll, can significantly correct handling properties by making the vehicle more or less understeering, it is also true that such corrections are delayed as compared to the maneuver starting time; the best rule is to design suspension mechanisms without any variation of toe angles and camber angles with the suspension stroke.

The desired level of understeering may be achieved by convenient tire sizing and also by optimizing weight distribution in consequence of transfers due to inertia forces.

This condition on toe angles is relatively easy to achieve; as far as camber is concerned the condition can be reached only if camber angles recover the roll angle effect, except for rigid axles where the problem does not exist.

This condition should not be assumed without consideration for the camber stiffness of selected tires.

What must be reduced is the additional cornering force rising from camber angles; if on a given tire camber stiffness is about 1/10 of the cornering stiffness (which means the cornering force arising from a side slip angle of 1 deg, with no camber, is identical to that arising from a camber angle of 10 deg, with no side slip angle), tires with a lower camber stiffness can be manufactured, making a simpler McPherson suspension perform like a more sophisticated double wishbone suspension.

In prevailing use conditions toe angles and camber angles must be set to zero to minimize wear and rolling resistance.

For the same reason it is better to minimize track variations because of the suspension stroke.

Comfort

Considering comfort, a first rule can be established for total suspension flexibility determined by coil springs and stop springs; a convenient value for the first natural frequency of the sprung mass should be set around 1 Hz, optimum for a human body in sitting position.

This condition implies that flexibility should be proportional to the vertical suspended mass or that suspension must become stiffer as the payload is increased.

The suspension total stroke must be limited for practical reasons: The wheel case dimension and the need to lift the car with a jack, in order to change a punctured wheel; it should be at least comparable with the extension of maximum road irregularities to be afforded. This condition has, as a consequence, a maximum stroke of about 70 mm in both direction (extension and compression) beginning with the reference condition of the vehicle at minimum load.

In the compression direction, the stroke necessary to reach the maximum payload condition must also be added. In conditions of wide load variations (station wagons, minivans, large sedans, etc.) suspensions with trim control are necessary.

To think about vertical obstacles is only a rough approximation of the reality. Because of vehicle motion each obstacle hits the wheel in a horizontal direction as well. It is fundamental to assure that each wheel also has a horizontal suspension; longitudinal stiffness should be no more then ten times the vertical.

This result may be obtained by designing a mechanism so as to have a convenient wheelbase variation with the suspension stroke; for front wheels, the wheelbase must decrease by the compression stroke, while the rear wheels must do the opposite. In addition elastic bushings must assure an additional convenient longitudinal flexibility.

In the first case the suspension is again a mechanism with one degree of freedom, but the path of the center of the tire patch has been conveniently

inclined in the vertical xz plane. In the second case an additional degree of freedom has been added: The longitudinal motion.

The suspension mechanism must never comply horizontally with toe angle variations, but in a reduced way for stabilizing braking.

Steering

Further considerations can be made for the steering mechanism.

Contrary to what is commonly thought, reaching the Ackermann condition is not vital until the difference between the effective and the theoretical angle (the so-called *steering error*) is lower than about 4 deg at maximum steering angle. In fact, side slip angles partially diminish steering angle accuracy.

The king-pin offset can affect steering wheel torque during braking; the rule for a negative king-pin offset is universally adopted, because it stabilizes the car path during braking (if the braking force increases on one wheel, the added steering action provided by the late response of the driver is able to counterbalance the yaw moment caused by the imbalanced braking forces).

The longitudinal trail, by cornering forces, can apply to the steering wheel an additional self-aligning torque; this contribution is particularly important because it can compensate for tire self-aligning torque decrease at large side slip angles, while still providing good cornering force.

It is better not to exaggerate with this practice by front wheel drive cars, because an aligning torque caused by traction force already exists.

The caster angle is bound to the longitudinal trail; an appropriate value can have a positive impact on the turning radius, because the king-pin inclination causes additional camber angles on the wheels.

Elasto-kinematic calculation

Independent suspensions

Elasto-kinematic behavior predictions can be made by modelling an articulated system including all spherical or cylindrical couples included in the mechanism. If the couples are considered to be stiff, the *kinematic* behavior can be predicted; if they are intended to be compliant to the applied forces, the *elasto-kinematic* behavior will be predicted.

It is convenient to calculate kinematic contributions to the suspension configuration first, adding to this configuration the effect of elastic displacements. Although this procedure is approximate, because it does not take into account the contributions of the modified position on the values of the forces; it is nevertheless acceptable in practical cases.

In these kinds of calculations, conventionally, all displacement and rotations of the wheel are referred to the car reference system and not to the wheel itself; in other words, the body is assumed to be fixed at a test bench block and the actual position of the wheel is calculated by imposing upon the hub a given displacement along the z axis.

At this displaced position a certain force F_z will be necessary to maintain the elastic suspension system in equilibrium; when necessary, to this new displaced position forces F_x and F_y will be applied to examine their additional effect.

It must be recalled that toe and camber angles that are so evaluated cannot be introduced into the formulae used to calculate tire forces, without being corrected by the effects of body and wheel rotations, such as roll angle, vehicle side slip angle and wheel steering angle.

Usually the study of the displacement and rotations caused by the suspension stroke are separated from the study of the effect of applied forces.

In the first case, vertical loads (at the wheel center or at the contact patch center), toe, camber, track, and wheelbase variations are calculated with reference to the static configuration, as a function of the suspension stroke (referred to the body) and for different steering angles.

These characteristics may have a different behavior in the case of symmetric stroke (the wheels of the same axle have the same stroke) or asymmetric (when the displacements are opposite, with the same absolute value).

In elasto-kinematic calculations, suspension elements and articulations to the body have further displacements because of the applied loads.

By calculating the effect of additional forces other than the vertical ones, the procedure is the same; it should be remembered that the path of the center of the contact patch, where forces are applied, is not perfectly directed along the z axis. The application of the F_x and F_y forces therefore causes additional suspension stroke.

At this new position, the effect of bearing and bushings deformations can be calculated.

In Fig. 3.73 some typical diagrams are shown that are obtained with this kind of analysis. The diagram at the upper left reports stroke as a function of the vertical force. The stroke Δz is measured beginning with the reference conditions of a car with full tanks and two people with their baggage; Δz is positive in compression and negative in extension.

Part of the line is almost linear, when only the coil spring is working (the contribution of the elastic bushings is very small), while at increasing loads the curvature of the diagram increases until it reaches a maximum value, as the stop spring starts to work. Suspension flexibility is defined as the partial derivative $\partial z/\partial F_z$, calculated for $\Delta z = 0$. The total suspension stroke is determined by the difference between the maximum compression and the maximum extension.

The diagram at the upper right shows toe angle as a function of the suspension stroke for different steering angles, while the diagram at the lower left concerns camber angle as a function of the suspension stroke.

The last diagram reports semi-track variation as a function of the suspension stroke.

These diagrams can also be used to calculate the same magnitudes as a function of roll angle by reading the diagrams at symmetric values, which easily provide the roll angle. It should be noticed that each of these diagrams as a func-

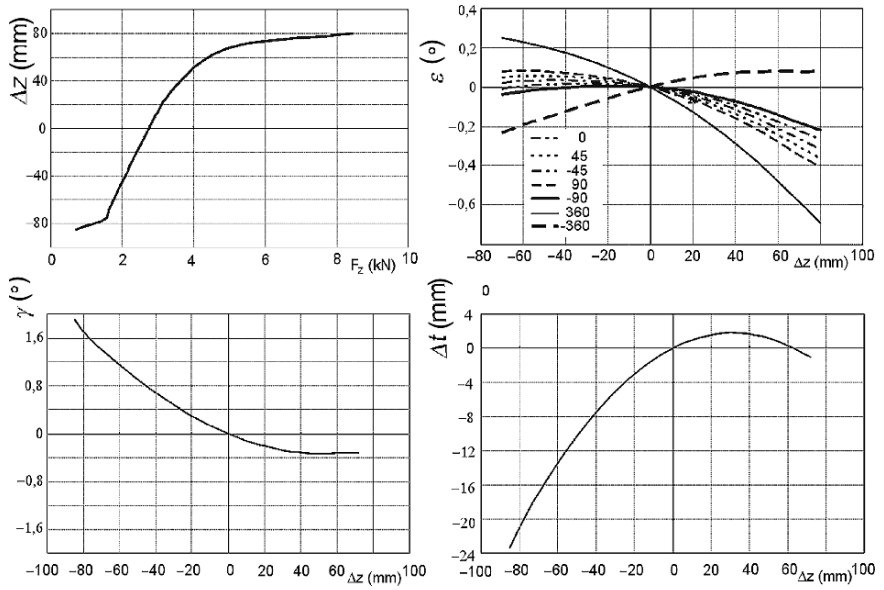


FIGURE 3.73. Typical diagrams of suspension geometric parameters. From the top right, moving clockwise, we can see the suspension stroke diagram as a function of the vertical load, the toe angle diagram as a function of the suspension stroke for different steering angles, the camber angle diagram as a function of the suspension stroke and the semi-track diagram as a function of the suspension stroke.

tion of roll is valid only for the value of the symmetric stroke we have considered as the starting point of the roll motion.

In the following Fig. 3.74 are represented the effects of the applied forces on the same characteristic parameters.

Toe angles variations as a function of the cornering force F_y, are shown in the upper left diagram. In this diagram the force is applied to a suspension only (called loaded side), but the effects are also measured on the other suspension of the same axle (called unloaded side): Notice how the unloaded side wheel responds, because of the mechanical connection established by the steering rack.

The following diagrams show the camber angle variation as a function of the cornering force F_y.

Toe angle variations as a function of the self-aligning torque M_z are shown at the lower left; in this case too the effect on the unloaded wheel can be observed.

The last two diagrams show respectively toe angle variation and wheelbase variation as a function of the braking or driving force F_x. This last diagram is useful for evaluating the longitudinal absorption capacity of longitudinal shocks.

These diagrams are often linearized by replacing these curves with their tangent at the origin, in order to allow input into simple mathematical models: in this case the derivatives of this parameter are considered.

The main derivatives are $(t)_{,z}$, $(t)_{,\phi}$, $(\gamma)_{,z}$, $(\gamma)_{,\phi}$ and $(\varepsilon)_{,z}$, $(\varepsilon)_{,\phi}$.

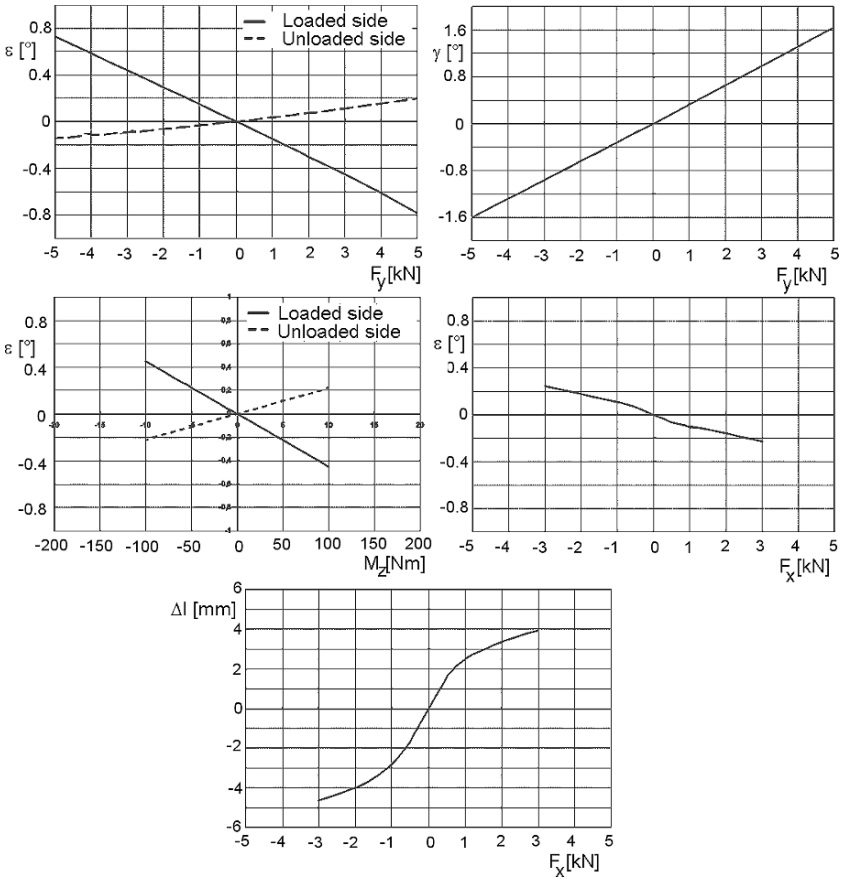


FIGURE 3.74. Typical diagrams describing the elasto-kinematic behavior of a suspension; notice how steering rack can couple the two wheels of the same axle.

Rigid axles

If the wheels are connected by a rigid beam, the latter must have two degrees of freedom with respect to the body, one translation in the z direction and one rotation around the roll axis. This rigid beam can be guided, as we have seen, by linkages or by leafs springs or by both.

The elasto-kinematic behavior must, in this case, be described by two coordinates; a common choice is to have diagrams as a function of roll angle, with suspension stroke as a parameter.

The solution with leaf springs, with structural function, has the disadvantage of approximating the correct kinematic behavior in which the axle can move only along the z coordinate with rotation about the roll axis poorly. The stiffness in the x and y directions, although far higher than that in the z direction, is not high enough, nor is the stiffness for rotations about the y axis. Above all, any

rolling motion is linked to a steering of the whole axis (roll steer). In other words, the derivative $\partial\delta/\partial\phi$ can be quite high (Fig. 3.63). The latter characteristic is due to the fact that the motion of the points at which the axle is connected to the springs is not exactly vertical, although the deviation from the vertical direction of the trajectory shown in the figure is exaggerated to illustrate the phenomenon.

The solution with guiding linkages can avoid this poor characteristic and also improve the behavior by longitudinal flexibility.

In all cases track and camber angle are constant, assuming tire vertical deformation is neglected. Therefore $(t)_{,z} = (t)_{,\phi} = (\gamma)_{,z} = (\gamma)_{,\phi} = 0$.

Roll axis

Roll axis is the instantaneous rotation axis of the vehicle when the body is in a symmetric condition (i.e.: $\phi = 0$). It is an instantaneous rotation axis because the roll motion is not a pure rotation around a physical articulation; therefore this axis can be defined for small rotations only, starting from a well defined initial condition.

The points where the roll axis crosses vertical planes through wheel centers of the same axles are called suspension *roll centers* CR.

For symmetry reasons, the roll axis must lie in the symmetry plane of the vehicle (xz plane) and therefore the roll centres of the suspensions must also lie in it. Note that in the case of a two-axle vehicle the roll centre of each suspension can be determined from the characteristics of the relevant suspension only, and that the roll axis can be defined as a line connecting the roll centres of the two suspensions. If the vehicle has more than two axles, the roll centres of the suspensions need not be aligned: A roll axis still exists, but it does not pass through the roll centres of the single suspensions, which are considered as insulated.

The roll centre of each suspension can also be defined as the point on a plane perpendicular to the ground and to the symmetry plane in which the application of a lateral force F_y to the vehicle body does not cause any roll. The two definitions obviously coincide.

Other important points are:

- The center of rotation of the body with respect to the wheels CW and
- The center of rotation of the wheels with respect to the ground RS

Rigid axles

Because in a rigid axle the two wheels are rigidly connected, the two points are located in the symmetry plane. Moreover, if the compliance of the tires is neglected, the wheels cannot rotate with respect to the ground: Points CW are located at infinity on the intersection between the ground and the plane parallel to the yz plane passing through the centres of the wheels. The two points BW coincide with the roll centre RC.

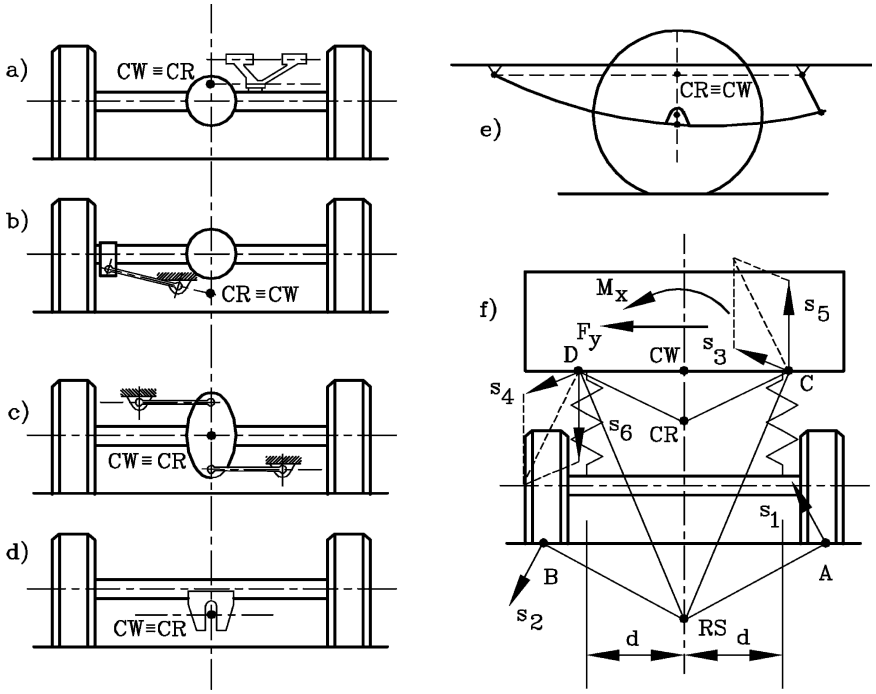


FIGURE 3.75. From a to e: position of point CW on various rigid axle suspensions; tires are stiff. Scheme f shows the position of RS, CR and CW, taking into account tire deformation.

If the compliance of the tires is accounted for, points RS lie on the symmetry plane slightly below the ground, but their positions are not exactly defined as they depend on the deflection of the tires and hence on the forces applied to them; it is, however, possible to define a zone in which they lie. Points CW coincide and are located in the symmetry plane of the vehicle below the roll centre CR.

In some case points CW are physically defined as there is a material hinge between the axle and the body (Fig. 3.75a–d). If the axle is guided laterally by leaf springs, CW is on the symmetry plane, at the level of the attachment of the springs to the body (Fig. 3.75e). The lateral deflection of springs causes CW to be located at a lower level and the inflection of the tires causes the roll centre CR to be located below CW.

With lateral deformations of tires, the center will be at the intersection of the perpendiculars to the displacement vectors, caused by the application of M_x and F_y, as in Fig. 3.75f.

A four-link suspension is shown in Fig. 3.76. To obtain the position of the roll centre, the intersections A and B of the axes of links 1-1′ and 2-2′ must be found first. These lie in the symmetry plane of the vehicle. The roll centre is found as the intersection of line AB with the plane perpendicular to the ground

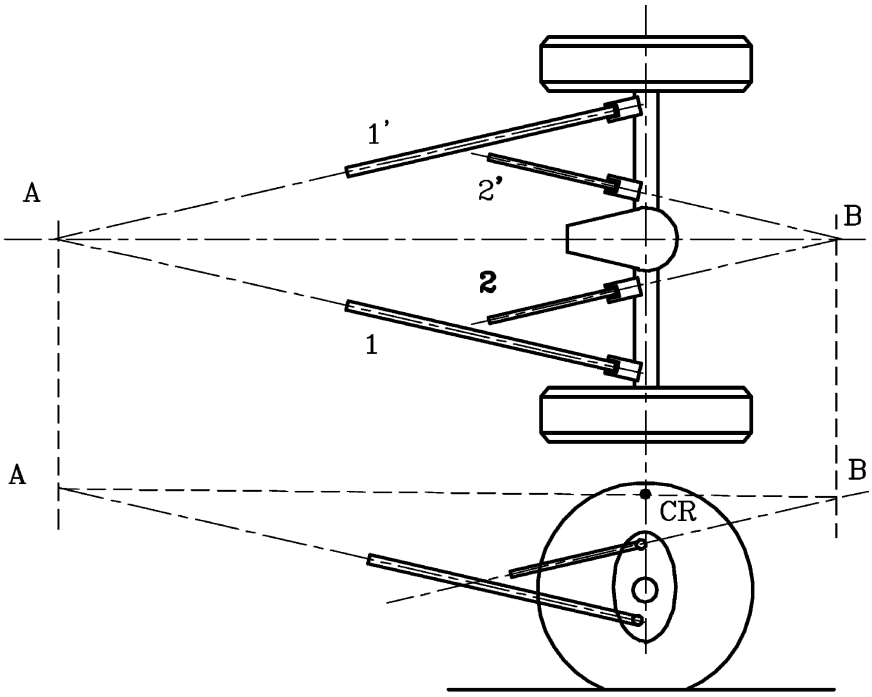


FIGURE 3.76. Four link suspension. Position of point CR.

containing the centres of the wheels. If two links are parallel (say links 1 and 1′) the intersection is at infinity and line AB is parallel to the projection of the relevant links on the symmetry plane.

The situation for a three-link suspension is similar (Fig. 3.77). The only difference is that point B is the intersection of the axis of the transversal link with the plane of symmetry.

In the case of a rigid axle with leaf springs, the roll centre is located at the intersection between the projection on the symmetry plane of a line connecting the points in which the springs are connected to the body and the perpendicular to the ground through the centre of the wheel (Fig. 3.75e). If the compliance of the tires is not neglected, to find an approximate location of CW, CR and RS it is possible to proceed as in Fig. 3.75f: A force F_y and a moment M_x are applied to the body: Displacements s_1, s_2, s_3 and s_4 of points A, B, C and D are due to the inflection of the tires, while the compliance of springs causes displacements s_5 and s_6, of points C and D. The intersection of the lines perpendicular to s_1 and s_2 locate point RS; CW is the intersection of the lines perpendicular to s_5 and s_6 and that of the perpendiculars to the resultants of s_3 and s_5 and of s_4 and s_6 locates point CR.

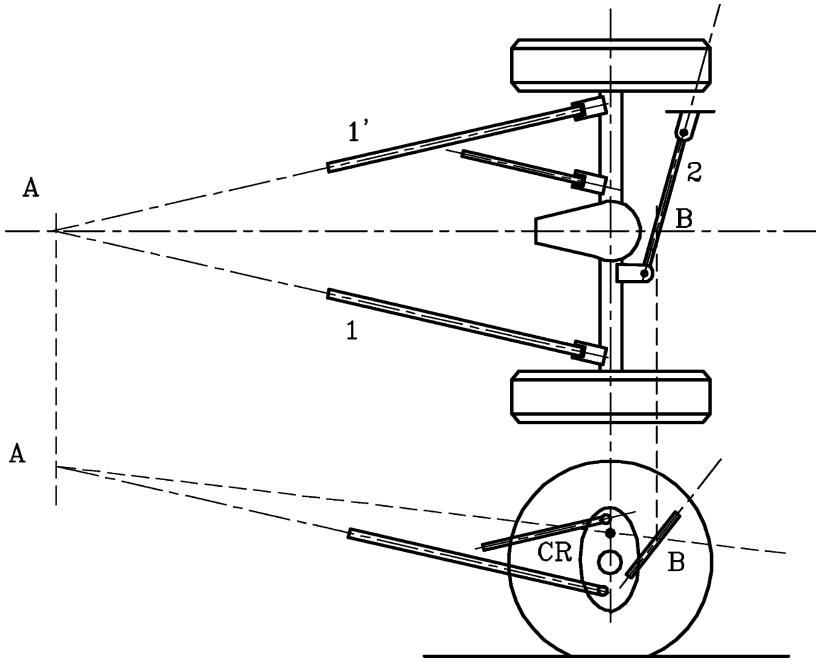


FIGURE 3.77. Three-link suspension. Position of point CR.

The derivative $(\gamma)_{,\phi}$ can be approximately computed as:

$$\frac{\partial \gamma}{\partial \phi} = \frac{\chi}{\chi + \Pi} \,, \tag{3.6}$$

where:

$$\chi = 2kd^2 + \chi_t \,,$$

(k is the stiffness of the springs and χ_t that of the anti-roll bar) and Π are respectively the rolling stiffness of the suspension and of the tires.

Independent suspensions

Note that in general the wheel motion is not planar. As a consequence the study of the kinematic behavior is not simple. It is, however, easy to obtain the exact kinematics of any suspension by using today's computer generated trajectories.

With double wishbone suspensions, the points CW1 and CW2 of the two wheels are located at the intersection of the upper arms and lower arms axis, which can cross outside of the vehicle (Fig. 3.79a) or inside (Fig. 3.78). It is possible to obtain $\partial t/\partial z = 0$ by demanding that CW1 and CW2 are on the ground (Fig. 3.79b), but this condition can be obtained only for a certain value of the pay load. If $\partial \gamma/\partial \phi$ must be set to zero, the points CW1, CW2 and CR must lie on the symmetry plane (Fig. 3.79c).

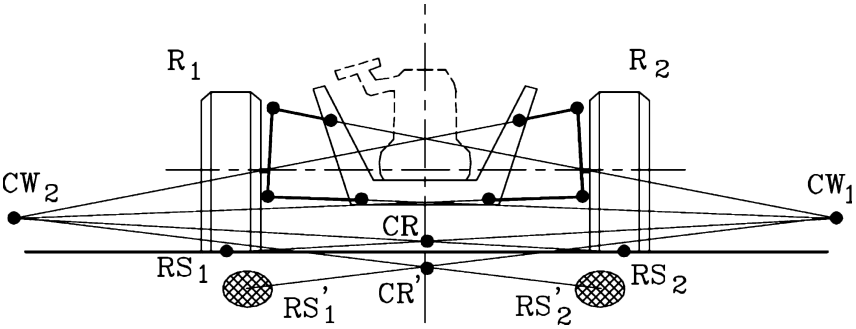


FIGURE 3.78. Position of roll center CR for a low double wishbone suspension with $1'2'$ and $3'4'$ arms parallel to the x axis. CR: position with rigid tires; CR′: position with deformable tires.

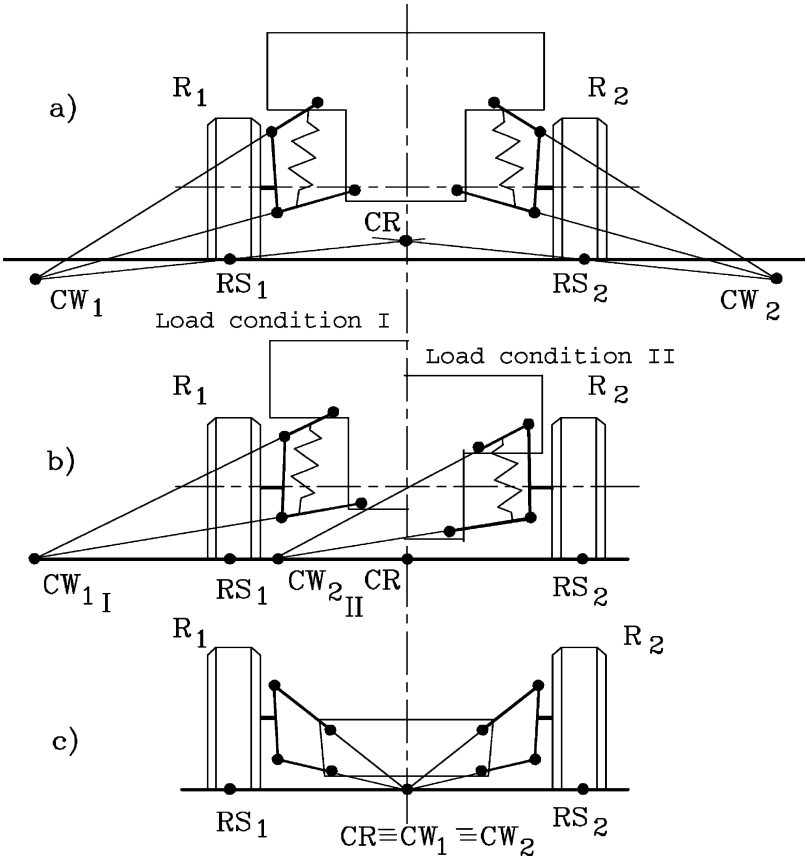


FIGURE 3.79. Low double wishbone suspensions.

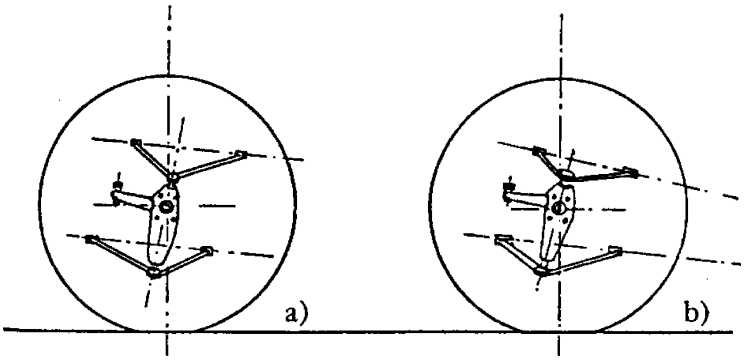


FIGURE 3.80. Low double wishbone suspension with non-horizontal articulation axis (a) and non-parallel axis (b). The inclination of these axis determines the position of the pitch center.

If the wheels were thin rigid discs, points RS1 and RS2 would coincide with the centres of the contact areas. If the compliance of tires is accounted for their approximated position can be located under the ground, slightly inboard of the centres of contact. By connecting points CW1 and RS1 and points CW2 and RS2 and intersecting such lines the roll centre CR, which lies in the symmetry plane, can be located. In the case of transversal articulated quadrilaterals, it is usually close to the ground or, if the deformation of the tires is considered, even below it. If the axes of the hinges of the two triangular linkages are not horizontal or parallel (Fig. 3.80) the determination of the roll centre and of the motion of the latter is far more complicated.

The construction necessary to obtain the different centers of a McPherson suspension is shown in Fig. 3.5.

A different approach is to use trailing arms (Fig. 3.44). The arms can be hinged to an axis that is perpendicular to the symmetry plane of the vehicle, but this is not always the case. In the first case the track remains constant,

$$\frac{\partial t}{\partial z} = \frac{\partial t}{\partial \phi} = 0$$

and the camber angle does not change in the vertical motions and is equal to the roll angle:

$$\frac{\partial \gamma}{\partial z} = 0 \ , \quad \frac{\partial \gamma}{\partial \phi} = 1 \ .$$

If the compliance of tires is neglected, the roll centre is on the ground (Fig. 3.44a) or slightly below (Fig. 3.44b).

Another solution is that based on swing arms. The hinges of the arms can be located at different points (Fig. 3.81a) or be coincident (Fig. 3.81b). The roll centre can be quite high on the road and the values of $\partial t/\partial z$ and of $\partial t/\partial \phi$ cannot be low. Swing arms can be connected to the power train block instead of

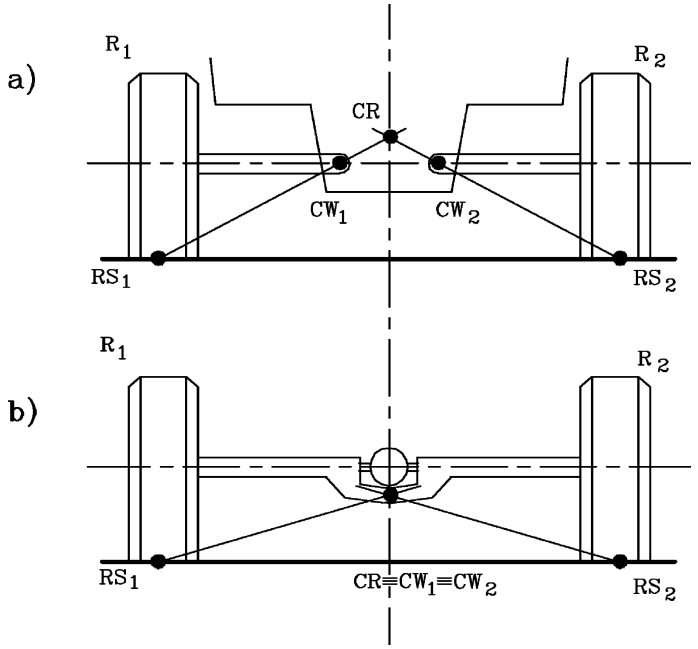


FIGURE 3.81. Swing arm suspension with different (a) or coincident (b) hinges.

being hinged to the body, as was common in small cars with rear engines. The fact that the engine was suspended using rubber blocks decreased the precision requirements of the suspension.

Conclusions

The primary kinematic characteristics of the various type of suspensions are summarized in Table 3.1. This overview does not cover all the solutions which have been used or suggested, such as, for example, the multilink solution, which is common on luxury cars. This is similar to a suspension with transversal quadrilaterals but the greater number of distinct elements allows a more detailed adjustment.

Pitch axis: anti-dive and anti-squat configurations

When the vehicle accelerates or brakes a load transfer between front and rear wheels occurs. This causes the body to pitch up (lift or squat) or down (dive).

In terms of roll, a pitch axis can be defined. The pitch axis is the axis of instantaneous rotation for pitch. It is an instantaneous rotation because the pitch motion is not a pure rotation around a well defined axis. This axis can be defined only for small rotations starting from a defined position, such as that where the x axis is parallel to the X axis of the inertial reference system.

TABLE 3.1. Primary characteristics of various suspensions. TA: Trailing arms; SAL: Swinging arms (limit case); SA: Swinging arms; DWP: Double wishbone with parallel arms; DWL: Double wishbone (limit case); DW: Double wishbone; MP: McPherson. A: CR high on the ground; B: below or on the ground; C: about at the wheel center; D: over the wheel center; E: on the ground.

	Rigid	Independent suspensions						
	axle	TA	SAL	SA	DWP	DWL	DW	MP
$\partial t/\partial z$	0	0	#0	#0	#0[a]	0	#0[b]	#0
$\partial t/\partial \phi$	0	0	0	#0	#0	#0	#0	#0
$\partial \gamma/\partial z$	0	0	#0	#0	#0	#0	#0	#0
$\partial \gamma/\partial \phi$	0	1	0	#0	1	#0	$>0<1$	$<0>1$
CR Pos.	A	B	C	D	B	E	B	B
CR Var.	0 ($\approx$0)	0	0	0	#0	0	#0	#0
$\partial i/\partial z$[c]	0	#0	#0	#0	0	0	#0	#0

[a] 0 in one configuration
[b] 0 in two configurations
[c] i: kingpin inclination

As forces F_{z_1} and F_{z_2} can apparently be approximated as:

$$\begin{cases} F_{z_1} = F^*_{z_1} - m\dfrac{h_G}{l}\dot{V} \\ F_{z_2} = F^*_{z_2} + m\dfrac{h_G}{l}\dot{V}\ , \end{cases} \tag{3.7}$$

where forces $F^*_{z_i}$ are those occurring when the vehicle does not accelerate, the lift of the front and the rear of the body are respectively:

$$\begin{cases} \Delta z_1 = m\dfrac{h_G}{lK_a}\dot{V} \\ \Delta z_2 = -m\dfrac{h_G}{lK_r}\dot{V}\ , \end{cases} \tag{3.8}$$

where K_f and K_r are the vertical stiffness of the front and rear suspensions. The pitch angle due to an acceleration is then:

$$\theta = \frac{1}{l}\left(-\Delta z_1 + \Delta z_2\right) = -m\frac{h_G}{l^2}\dot{V}\left(\frac{1}{K_a} + \frac{1}{K_r}\right). \tag{3.9}$$

A positive value of θ occurs when the vehicle dives (pitches down), as occurs with negative acceleration, hence the minus sign in the formula.

This expression is, however, an oversimplification, for two reasons: First the longitudinal forces due to the driving or braking wheels can themselves cause a pitching moment due to the coupling of the suspensions and, secondly, driving and braking torque reactions can be applied, at least partly, to the suspensions

instead of the body, inducing further pitching. Both effects cause pitching even in constant speed driving.

If suspensions allow the wheels to move in the x direction, i.e. if the characteristic $\partial x/\partial z$ is not vanishingly small, a fraction $F_x \partial x/\partial z$ of force F_x acting between the road and the wheel acts on the suspension and causes pitching. Equation (3.7) then becomes:

$$\begin{cases} F_{z_1} = F^*_{z_1} - m\dfrac{h_G}{l}\dot{V} - \left(\dfrac{\partial x}{\partial z}\right)_1 F_{x1} \\ F_{z_2} = F^*_{z_2} + m\dfrac{h_G}{l}\dot{V} - \left(\dfrac{\partial x}{\partial z}\right)_2 F_{x2} \end{cases} \tag{3.10}$$

and Eq. (3.9) transforms into:

$$\theta = -m\frac{h_G}{l^2}\dot{V}\left(\frac{1}{K_a} + \frac{1}{K_r}\right) - \left(\frac{\partial x}{\partial z}\right)_1 \frac{F_{x1}}{lK_a} + \left(\frac{\partial x}{\partial z}\right)_2 \frac{F_{x2}}{lK_r} \,. \tag{3.11}$$

If only longitudinal forces needed to accelerate or brake the vehicle are considered and the percentage of the longitudinal force assigned to the front axle is k_l, it follows that:

$$F_{x_1} = k_l m\dot{V} \,, \qquad F_{x_2} = (1 - k_l)m\dot{V} \,.$$

Equation (3.9) yields:

$$\theta = -m\frac{\dot{V}}{l}\left[\frac{h_G}{lK_a} + \frac{h_G}{lK_r} + \frac{k_l}{K_a}\left(\frac{\partial x}{\partial z}\right)_1 - \frac{(1-k_l)}{K_r}\left(\frac{\partial x}{\partial z}\right)_2\right] \,. \tag{3.12}$$

Obviously Eq. (3.9) holds in the case of acceleration and braking alike, provided that the sign of $\dot{V}$ is correct and a suitable value for k_l is used.

Consider, for example, the trailing arm suspension of Fig. 3.82a. With simple geometrical reasoning it is easy to assess that:

$$\left(\frac{\partial x}{\partial z}\right) = \frac{e}{d} \,. \tag{3.13}$$

A similar equation also holds for the suspension of Fig. 3.82b. If a torque M_y is applied to the sprung mass, it causes an increase of the force acting on the spring equal to M_y/d. As the torque linked to the generation of driving or braking forces is equal to $-F_x R_l$, the result is that the application of the braking torque to the suspension can be accounted for by substituting:

$$\left(\frac{\partial x}{\partial z}\right)_i + \left(\frac{R_l}{d}\right)_i \qquad \text{to:} \qquad \left(\frac{\partial x}{\partial z}\right)_i \,.$$

Note that d is positive when point A is in front of the wheel and negative otherwise.

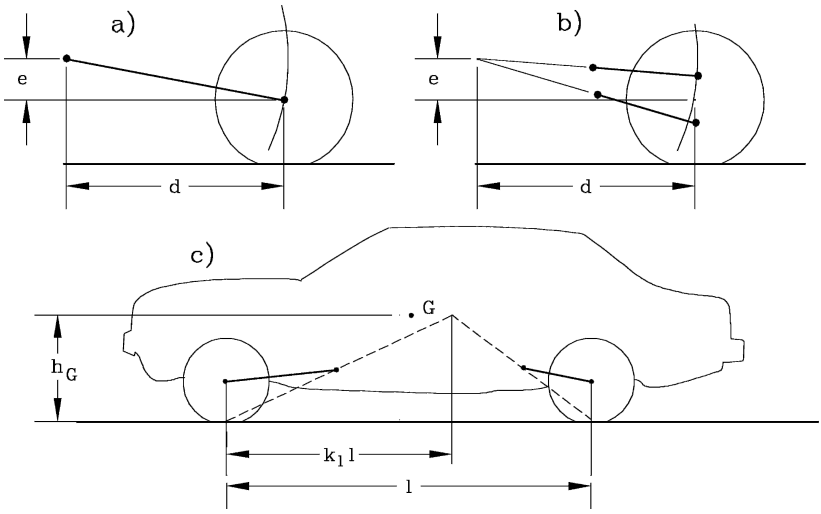


FIGURE 3.82. Relationship between $\partial x/\partial z$ and suspension geometry, for trailing arms (a) and double wishbone (b) suspensions. If arms are parallel, $d \to \infty$. How to establish the anti-dive effect (c).

The driving torque is applied to the unsprung mass in the case of live axles, while in De Dion axles and independent suspensions it is applied directly to the vehicle body and this correction does not apply. Braking torques are, on the contrary, usually applied to the unsprung masses, so the term in R_l/d must always be accounted for. However, if the torque transmission between the sprung and the unsprung masses is supplied by linkages which prevent any relative rotation about the y axis, these effects are minimized as d tends to infinity.

The above relationships allow one to design the suspensions to compensate, usually partially, for squat or dive. A total compensation occurs when Eq. (3.11) yields $\theta = 0$. If:

$$\frac{h_G}{lK_a} + \frac{k_l}{K_a}\left(\frac{\partial x}{\partial z}\right)_1 = 0 \tag{3.14}$$

the front of the car does not lift in acceleration or dive in braking, while if:

$$\frac{h_G}{lK_r} - \frac{(1-k_l)}{K_r}\left(\frac{\partial x}{\partial z}\right)_2 = 0 \tag{3.15}$$

the rear does not squat in acceleration or lift in braking.

Note that in case of a single driving axle either $k_l = 0$ or $k_l = 1$ and both front and rear compensations cannot be performed together. To obtain a complete compensation the term in square brackets in Eq. (3.11) must vanish and the front of the car must dive to compensate for the squat of the rear axle in front drives. The rear must lift in rear drives.

In case of braking a total compensation of the front axle leads to the condition:

$$\frac{k_l}{K_a}\left[\left(\frac{\partial x}{\partial z}\right)_1+\left(\frac{R_l}{d}\right)_1\right]=-\frac{h_G}{lK_a}\,, \tag{3.16}$$

i.e.:

$$\frac{k_l}{K_a}\left(\frac{e+R_l}{d}\right)_1=-\frac{h_G}{lK_a} \tag{3.17}$$

and that of the rear axle to:

$$\frac{(1-k_l)}{K_r}\left[\left(\frac{\partial x}{\partial z}\right)_2+\left(\frac{R_l}{d}\right)_2\right]=\frac{h_G}{lK_r}\,, \tag{3.18}$$

i.e.:

$$\frac{(1-k_l)}{K_r}\left(\frac{e+R_l}{d}\right)_2=\frac{h_G}{lK_r}\,. \tag{3.19}$$

A simple geometrical construction is shown in Fig. 3.82c; if the pivots are on the dotted lines, a complete compensation in braking is obtained while if they lie below the lines dive compensation is only partial. If they are above the line, the front will lift and the rear will squat in braking.

A similar scheme could be advanced for traction conditions; in this case the dotted line should start at the wheel center and not on the contact point.

Usually no complete anti-dive compensation is obtained for many reasons, some of them psychological (a flat braking is not desirable) and some objective (complete anti-dive compensation can lead to overcompensation of acceleration squat and to geometries which are poor both for comfort and performance).

3.6.4 Bench testing methods

The most relevant problems of testing a suspension are those that are common to the development of the remaining vehicle components. These are:

- Reducing lead time for significant prototypes
- Improving speed of performing tests
- Containing testing costs

For these reasons it is preferable to test suspensions as far from the actual road as possible, using dedicated indoor benches; obviously the final approval will only be given after an on road test of the complete vehicle.

The advantage of this approach is bound to the simplification of the system to be analyzed.

- The test object can be simplified, by limiting prototypes to the genuinely new components and by utilizing remaining ones taken from the shelf of the production shop. This approach can be used for both a suspension to be bench tested and a vehicle, drivable on the road with the new suspensions. Vehicle of this kind are usually called *mule cars*.

- Test outputs can be easily obtained in an objective way for prototype simplicity and because testing is not performed on the road; these outputs can be elaborated through mathematical models to predict the performance of the new vehicle.
- Bench tests can be easily automated, with benefits for costs and the confidence level of the test result.

This approach is clearly advantageous only if the test results have been anticipated by calculations; in this case only the most relevant issues are investigated, while those less relevant can be ignored.

Elasto-kinematic bench

Bench tests finalized to elasto-kinematic behavior measurement are performed essentially in order to:

- Verify the accomplishment of the projected characteristics
- Determine the input data of vehicle dynamics simulation models, used to analyze and verify the performance of a prototype car or interpret the performance of a competitor's car (*reverse engineering*)

Generally speaking an elasto-kinematic bench is a system able to impose upon the suspensions of a real car or a test fixture, which is fixed on the bench, an assigned displacement or set of forces.

The bench is composed of:

- Actuators (can be hydraulic or electric)
- A control system for the actuators
- A measurement and recording system
- A suitable software package for data acquisition, elaboration and hardware control

As an example, an elasto-kinematic bench with hydraulic actuators and anthropomorphic arm measuring system is briefly described.

Actuators are composed of a pair of moving platforms, shown in Fig. 3.83, able to apply assigned displacement and forces to the center of the contact patch of the tire or to the wheel hub.

Displacements can be parallel or opposed, along the three axes; rotations can also be imposed around the z axis; both displacement and rotations are applied according to sinusoidal laws at very low frequency ($\approx$0.005 Hz) within the desired measuring range.

The possibility of applying load cycles is essential for evaluating frictions and hystereses.

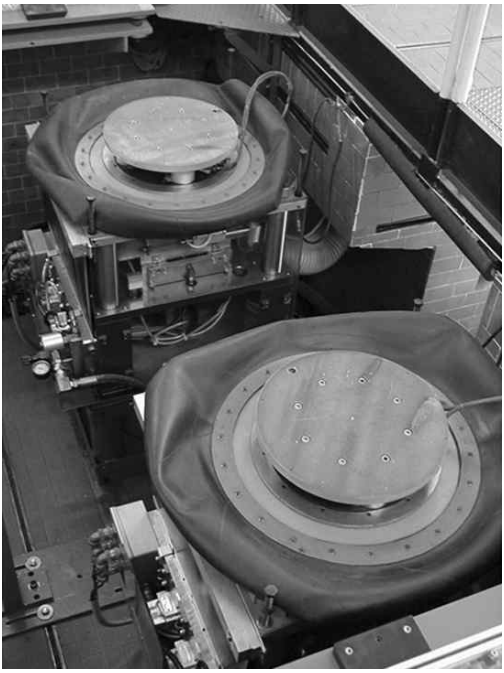


FIGURE 3.83. Elasto-kinematic bench with hydraulic actuators.

Assigned vertical or cornering forces and self-aligning torque can also be applied.

The anthropomorphic arm system of Fig. 3.84 allows the spatial motion of the wheel hubs of the same axle to be measured, consequent to the application of the said displacements and forces; three displacements and three angles are monitored.

The body is fixed through the structure 1 and is considered as the reference system; measurements can be repeated with different steering angles.

While interpreting measurements body displacements must also be evaluated. When assessing the suspension only, their contribution should be ignored; they are to be taken into account, however, when the behavior of the entire car is under investigation.

We conclude this example by reporting some of the technical specifications of this bench:

Vertical stroke:	$\pm$150 mm
Longitudinal displacement:	$\pm$75 mm
Lateral displacement:	$\pm$75 mm
Angle of rotation around Z':	$\pm$90 deg
Steering angle rotation:	$\pm$180 deg
Vertical force:	$\pm$20,000 N

FIGURE 3.84. Elasto-kinematic bench with hydraulic actuators and anthropomorphic arm measuring system. The anthropomorphic arm detects the spatial wheel motion, consequent to force application or suspension stroke.

Longitudinal force:	±10,000 N
Lateral force:	±10,000 N
Torque around Z':	±1,000 Nm
Steering wheel torque:	±50 Nm

Vibration characterization bench

This bench is also called a four axes bench, because of the number of available independent actuators; it allows measurements of the vibrational behavior of the entire vehicle and its suspensions caused by the application of periodic forces at its wheels, which simulate, in a controlled environment, the road profile.

Each wheel rests on an independent vertical actuator. This actuator is characterized by the following limits:

- Input forces can only be vertical; horizontal components due to car motion are neglected
- Wheels do not rotate; dynamic contribution produced on tires by rolling is, therefore, neglected.

The bench is, nevertheless, useful for monitoring the filtering capacity of the suspension, for comparing different suspensions, for reproducing and solving

problems detected on the road, in terms of vibration, noise and squeaks and for supplying experimental correlations to mathematical models on vehicle comfort.

The bench includes:

- Four vertical actuators
- Actuation control system
- Measurement system
- Software package for data acquisition and control

Different kinds of vertical inputs are available:

- Sinusoidal input at constant frequency
- Sinusoidal input at increasing frequency (frequency *sweep*)
- *Random* non-correlated inputs on the four wheels
- Assigned road profiles reproduction

The first three input types allow the transmissibility of the tire-suspension-body system to be calculated, as ratio between the input point at the wheel and the measurement point. By changing excitation amplitude nonlinearity can be discovered. Sinusoidal inputs and frequency sweeps facilitate their detection.

Non-correlated random inputs are useful for a quick characterization: With a single test a predefined frequency field my be investigated. By data elaboration it is possible to separate the transmissibility contributions of parallel springing, roll and pitch motions.

The reproduction of road profiles is a direct evaluation of the suspension comfort performance.

To characterize suspension comfort, three axial accelerations are measured at wheel hubs, body spring seats and shock absorber body connections.

The comparison of the accelerations of the body at foot rest, seat rails and steering wheels is a common indicator for comfort performance.

If the bench is not noisy induced noise can also be evaluated.

The technical specifications of a bench of this kind may be:

Suspension stroke:	± 100 mm
Maximum actuation speed:	2.8 m/s
Maximum acceleration:	270 m/s^2
Passing band:	up to 200 Hz.

Fatigue bench

Fatigue testing can be applied to the entire suspension or to a suspension component (like a suspension arm or a spring, etc.), separated from the system.

The benches in use depend on the complexity of the system to be tested. Components are tested using a hydraulic pulsator that can be combined with

other pulsators, according to the complexity of the system; a single hydraulic actuator should have these specifications:

- Three axial capability (can apply forces along the x, y, z axes simultaneously)
- Load programmer
- Complete body or universal fixation structure
- Signal recorder

The suspension to be tested is mounted on the body or body dummy, set on the bench with a combination of actuators; the same bench can also fatigue test an entire car body.

Figure 3.85 shows a bench of this kind: Suspensions are installed on the body, which is fixed to the bench through the structure 1. Suspensions are loaded by a set of actuators that can apply vertical, longitudinal and cornering forces.

Longitudinal and cornering force actuators work through rods 3 and 4.

Actuators must be designed for an unlimited life.

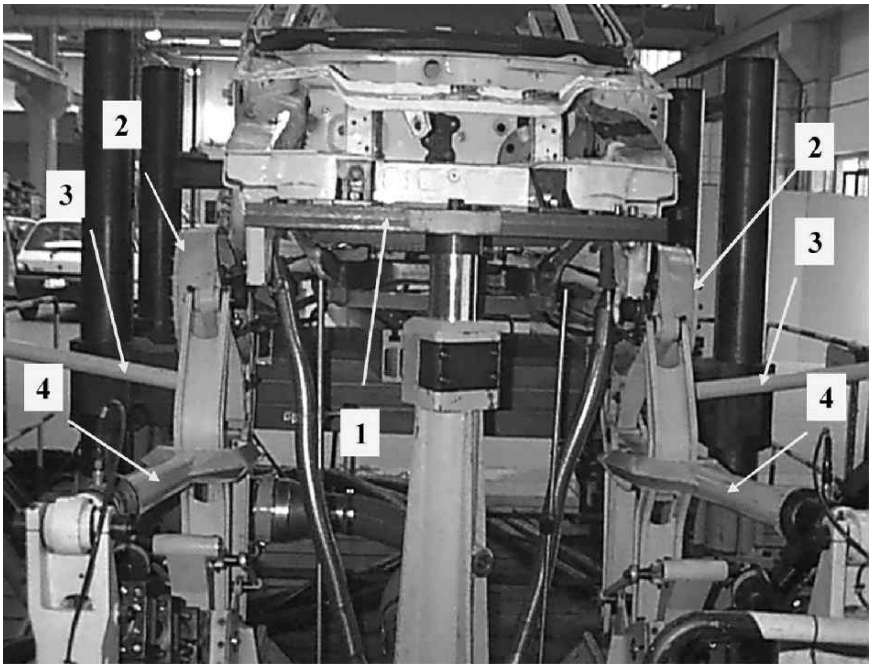


FIGURE 3.85. Three axial bench: the suspensions are mounted on a body, fixed to the bench with the structure 1. Suspensions are tested by actuators that can transfer to the wheel longitudinal, vertical and cornering forces.

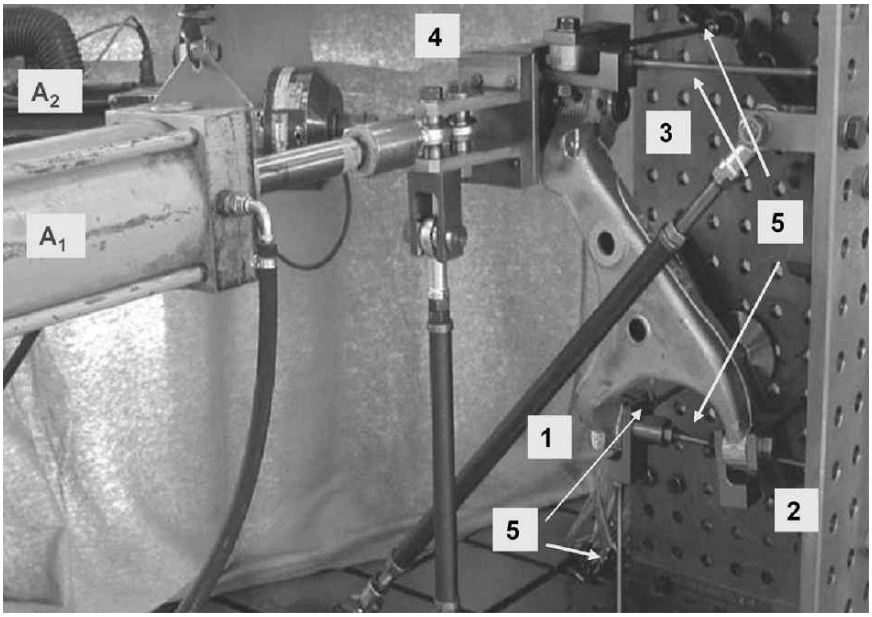


FIGURE 3.86. Fatigue test of a single suspension arm.

These actuators apply load blocks which are based on a load time history gathered on real roads; to shorten the test duration, some parts of this history are eliminated if they do not contribute to structural damage.

When a single component is tested, the bench is much smaller and can be used for different tests: It consists of a general purpose fixture for assembling the test components and the actuators and some hydraulic actuators.

These benches can be easily retooled to perform tests on different components.

Figure 3.86 represents the bench setup for testing a longitudinal arm of a guided arms suspension. The point 1 corresponds to the body articulation, points 2 and 3 to the other arm articulations; point 4 is the wheel hub flange.

The arm is installed on a test bed with thin rods 5, able to react with axial forces only and set so as to reproduce the total mechanism, which is missing on the bench. In particular, point 1 is fixed by three rods, as to eliminate any linear displacement, point 2 has the only vertical displacement blocked, while point 3 has the lateral and vertical displacement constricted.

Point 4, corresponding to the wheel hub flange, is loaded by two hydraulic actuators; the vertical actuator A_1 simulates the vertical static load, while a second actuator A_2 applies a periodic load on the yz plane, with a given inclination so as to reproduce the cornering force and the consequent load transfer.

A different setup could simulate traction and braking forces.

4
STEERING SYSTEM

4.1 INTRODUCTION

Vehicles may be classified into two categories, according to how their path is controlled:

- *Guided vehicles*, or better, *kinematically guided vehicles*, whose trajectory is fixed by a set of kinematic constraints.
- *Piloted vehicles*, in which the trajectory, a tri-dimensional or planar curve, is determined by a guidance system controlled by a human pilot or by a device, usually electro-mechanical. The guidance system acts by exerting forces on the vehicle that are able to change its trajectory.

In the first case the kinematic constraint exerts all forces needed to modify the trajectory without any deformation, i.e. it is assumed to be infinitely stiff and infinitely strong. Perfect kinematic guidance is therefore an abstraction, although it is well approximated in many actual cases.

In the second case the forces are due to changes in the attitude of the vehicle, which in turn are caused by forces and moments produced by the guidance devices. These vehicles can be said to be dynamically guided.

Apart from cases in which the forces needed to change the trajectory are directly exerted by thrusters (usually rockets), there can be two situations:

- The attitude changes can be quite large, large enough to be directly felt by the pilot or driver.
- The attitude changes can be small enough to be unnoticed.

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

The first case is that of aerodynamically or hydrodynamically controlled vehicles, in which the pilot acts on a control surface, causing the changes of attitude needed to generate the forces that modify the trajectory. There is usually a certain delay between the changes of attitude and the actual generation of forces. Consequently, the driver feels clearly that a dynamic control, i.e. a control through the application of forces, takes place.

In road vehicles the situation is similar but the driver has a completely different impression: The driver operates the steering wheel, causing some wheels to work with a sideslip and to generate lateral forces. These forces cause a change in the attitude of the vehicle (change of angle β) and then a sideslip of all wheels: The resulting forces alter the trajectory.

However, the linearity of the behavior of the tire and the high value of the cornering stiffness give the driver the impression of kinematic, not dynamic, driving. Wheels seem to be rolling unimpeded and the trajectory seems to be determined by the direction of the mid-planes of the wheels.

This impression has influenced the study of the handling of motor vehicles for a long time, originating the very concept of *kinematic steering* and in a sense obscuring the true meaning of the phenomena.

The impressions of the driver are in accord with this kinematic approach, at least in terms of the linear aspects of the tire's behavior. When high values of the sideslip angles are reached, the average driver has the impression of losing control of the vehicle, much more so if this occurs abruptly. This impression is confirmed by the fact that in normal road conditions, particularly if radial tires are used, the sideslip angles become large only when approaching the limit lateral forces.

These considerations are only an indication, as there are intermediate cases such as kinematic guidance with deformable constraints or magnetic levitation vehicles. The difference turns out to be more quantitative than qualitative, and depends largely on the greater or lesser stiffness with which the vehicle responds to variations of attitude due to the guidance devices.

We should also notice that it is conceptually possible to generate the torque, which initially modifies the vehicle attitude angle, by means other than those we have explained. Instead of steering and applying a side slip to the steering wheels, an unequal traction or braking force can be applied to the wheels of the same axle, or an aerodynamic force can be applied outside the vehicle symmetry plane.

In both cases, this torque causes a side slip angle on the wheels that curves the vehicle path. We could imagine a vehicle where no axle steers, one that is controlled by means of differential wheel braking. Tracked vehicles are controlled in a similar way, but their dynamics is quite different because of the intrinsic difference between tires and tracks.

As we will see later, differential braking is used in some control systems, applied to improve the vehicle's dynamic behavior or handling: The path is set up by the driver, who steers the wheels, while an automatic control system applies corrections by differential braking or, sometimes, through additional steering angles.

In the following paragraphs we will comment on the main components of the steering system, including:

- *Steering mechanism*, the system of linkages steering the front wheels in a particular way around the king-pin axis, connecting steering arms moving with the suspension stroke to the steering box
- *Steering box*, transforming steering wheel rotation into a displacement of the steering tie rods
- *Steering column*, connecting steering wheel with steering box

4.2 STEERING MECHANISM

Before explaining the steering mechanism configuration, let us introduce some further consideration on steering piloted road vehicles.

Low speed or kinematic steering is defined as the motion of a wheeled vehicle as determined by pure rolling of the wheels. The velocities of the centres of all the wheels lie in their mid-plane, i.e., the sideslip angles α are vanishingly small. In these conditions the wheels can exert no cornering force to balance the centrifugal force due to the curvature of the trajectory. Kinematic steering is possible only if the velocity is vanishingly small.

Consider a vehicle with four wheels, two of which can steer (Fig. 4.1). The relationship that must exist to allow kinematic steering is easily found, by imposing the condition that the perpendiculars to the mid-planes of the front wheels must meet those of the rear wheels at the same point:

$$\tan(\delta_1) = \frac{l}{R_1 - \dfrac{t}{2}}\,, \qquad \tan(\delta_2) = \frac{l}{R_1 + \dfrac{t}{2}}\,. \tag{4.1}$$

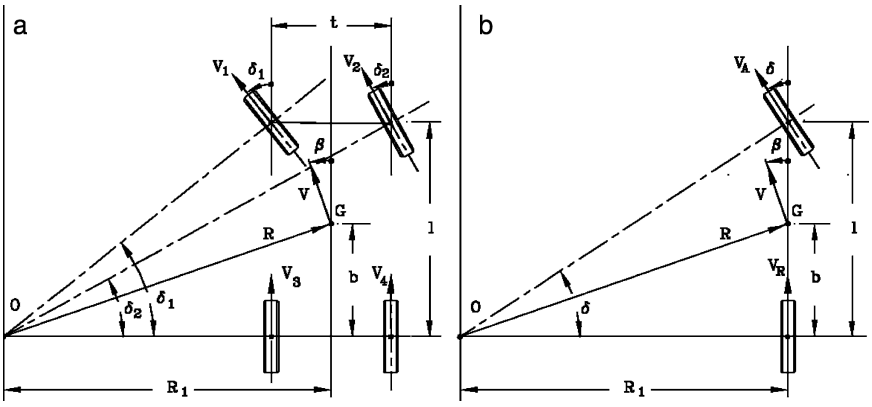


FIGURE 4.1. Kinematic steering of a four-wheeled and a two-wheeled vehicle.

Instead of the track t, Eq. (4.1) should contain the distance between the kingpin axes of the wheels, or better, between their intersections with the ground.

By eliminating R_1 between the two equations, a direct relationship between δ_1 and δ_2 is readily found:

$$\cot(\delta_1) - \cot(\delta_2) = \frac{t}{l} \; . \tag{4.2}$$

A device allowing wheels to be steered in exact accordance with Eq. (4.2), is usually referred to as *Ackerman steering* or *Ackerman geometry*. No actual steering mechanism allows such law to be followed exactly and a *steering error* $\Delta\delta_2$, defined as the difference between the actual value of δ_2 and that obtained from Eq. (4.2), can be obtained as a function of δ_1.

Consider, for example, the device based on an articulated quadrilateral shown in Fig. 4.2, used by rigid axle suspensions with a screw and sector steering box.

By this simple mechanism, the track rod AB coordinates wheel steering.

To calculate the relationship between δ_1 and δ_2, it is sufficient to develop some geometrical considerations. The track rod length AB is:

$$\left(\overline{\mathrm{A} - \mathrm{B}}\right) = l_1 - 2l_2 \sin(\gamma) \; . \tag{4.3}$$

When the steering wheel is turned, the new positions of A and B, referred to as A' and B', with reference to a reference system having its origin in O_1 and the x axis in the cross direction, are:

$$\left(\overline{\mathrm{A'} - \mathrm{O_1}}\right) = \left\{ \begin{array}{c} l_2 \sin(\gamma + \delta_1) \\ -l_2 \cos(\gamma + \delta_1) \end{array} \right\} \; , \tag{4.4}$$

$$\left(\overline{\mathrm{B'} - \mathrm{O_1}}\right) = \left\{ \begin{array}{c} l_1 - l_2 \sin(\gamma - \delta_2) \\ -l_2 \cos(\gamma - \delta_2) \end{array} \right\} \; , \tag{4.5}$$

The square of the distance between A' and B' is:

$$\begin{aligned} \left(\overline{\mathrm{A'} - \mathrm{B'}}\right)^2 \; &= \; [l_2 \sin(\gamma + \delta_1) + l_2 \sin(\gamma - \delta_2) - l_1]^2 + \\ &\quad [l_2 \cos(\gamma + \delta_1) - l_2 \cos(\gamma - \delta_2)]^2 \; . \end{aligned} \tag{4.6}$$

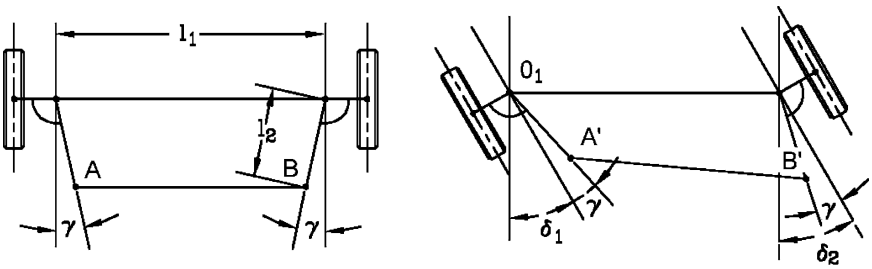


FIGURE 4.2. Scheme of a steering mechanism made by an articulated quadrilateral.

But this distance is equal to the square of $\left(\overline{\mathrm{A-B}}\right)$, and a relation between δ_1 and δ_2 can be obtained:

$$\sin(\gamma-\delta_2)+\sin(\gamma+\delta_1)=$$

$$=\frac{l_1}{l_2}-\sqrt{\left[\frac{l_1}{l_2}-2\sin(\gamma)\right]^2-\left[\cos(\gamma-\delta_2)-\cos(\gamma+\delta_1)\right]^2}\,, \tag{4.7}$$

that is:

$$1+\sin(\gamma-\delta_2)\sin(\gamma+\delta_1)-\lambda\sin(\gamma-\delta_2)-\lambda\sin(\gamma+\delta_1)+ \tag{4.8}$$

$$+\left[\lambda-2\sin(\gamma)\right]\sin(\gamma)-\cos(\gamma-\delta_2)\cos(\gamma+\delta_1)=0\,,$$

where $\lambda=l_1/l_2$.

Equation (4.7) may be solved so as to obtain δ_2 as a function of δ_1. With laborious calculations, the following relationship is obtained:

$$A\sin^2(\gamma-\delta_2)+B\sin(\gamma-\delta_2)+C=0\,,$$

where:

$$A=1+\lambda^2-2\sin(\gamma+\delta_1)$$

$$B=2D\left[\sin(\gamma+\delta_1)-\alpha\right]$$

$$C=D^2-\cos^2(\gamma-\delta_1)$$

$$D=1-\lambda\sin(\gamma+\delta_1)+\left[\alpha-2\sin(\gamma)\right]\sin(\gamma)\,,$$

which allows δ_2 to be calculated.

If the two steering arms converge on the rear wheel axle at the intersection with the vehicle symmetry plane, as proposed by Jeantaud, the angle γ can be calculated as:

$$\gamma=\mathrm{artg}\left(\frac{l_1}{2l}\right).$$

The relationship $\delta_2\,(\delta_1)$ obtained with the Jeantaud quadrilateral is compared with that obtained using the correct kinematic relationship and three different values of γ in Fig. 4.3a. The steering error $\Delta\delta_2=\delta_2-\delta_{2_c}$ can be calculated by any value of δ_1, as shown on Fig. 4.3b.

The Jeantaud condition leads to a value of $\gamma=13.3$ deg and fairly large steering errors. Three other curves are plotted for γ: 16 deg, 18 deg and 20 deg (in all three cases the steering arms converge behind the rear axle); the error decreases, for low values of the steering angle, as γ decreases. Nevertheless low errors at low steering angles are accompanied by large errors at large steering angles. A compromise is necessary: In our case a value of $\gamma=18$ deg can be a reasonable choice.

The complete calculation of the steering error for a rack and pinion steering box is omitted for sake of brevity; we refer to Fig. 4.4.

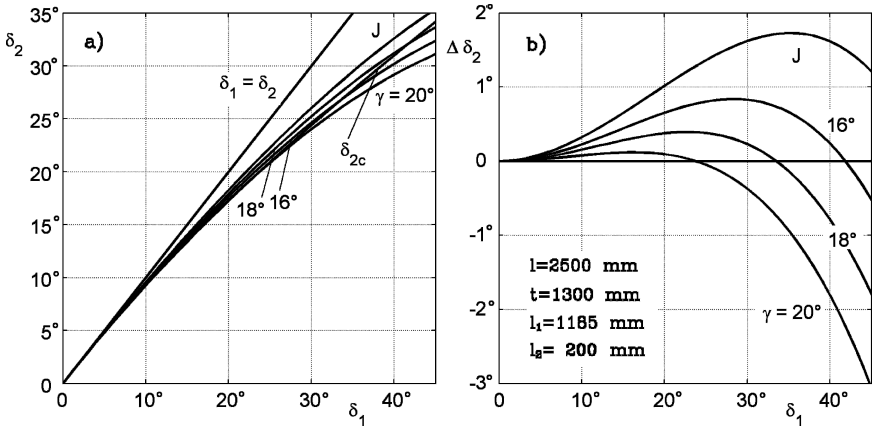


FIGURE 4.3. (a) Relationship $\delta_2\,(\delta_1)$ obtained with the Jeantaud quadrilateral (curve J) compared with that obtained using the correct kinematic relationship and three different values of γ. (b) Steering error, $\Delta\delta_2 = \delta_2 - \delta_{2_c}$ as a function of δ_1 for the same steering systems.

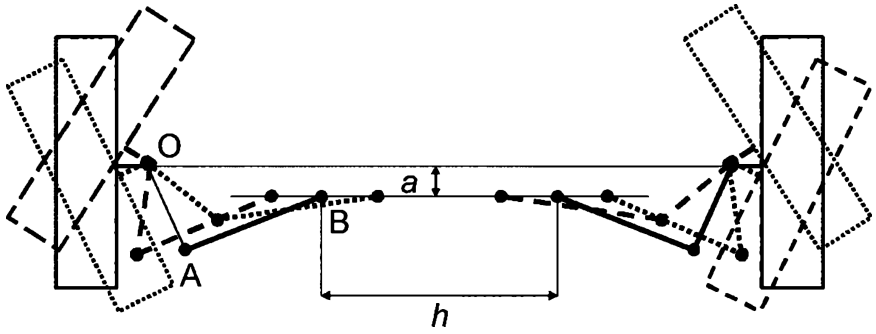


FIGURE 4.4. Kinematic scheme to calculate steering error by a rack and pinion steering box.

If we suppose that rack and tie rods OA lie on the same plane (this is not always true) or the triangle OAB is always on a horizontal plane, we can understand visually that the steering error depends on dimensions a and h, which are conditioned by the space allowable in the engine compartment.

Let us consider, finally, the case of a screw and sector steering box in connection with an independent wheel suspension; in this case a track rod cannot be applied, because it could introduce undesirable toe angle variations by asymmetric strokes of the suspensions.

Different steering mechanisms can be applied; some of these are shown in Fig. 4.5.

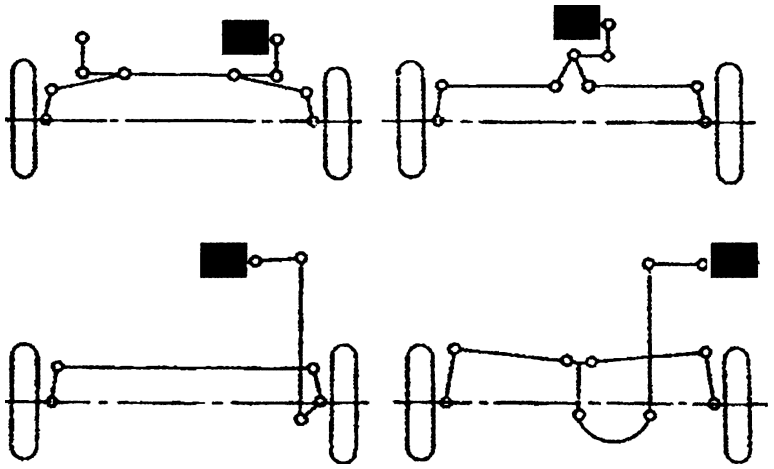


FIGURE 4.5. Steering mechanisms for a screw and sector steering box, applied to an independent wheel suspension; on the upper side of the figure, the case of a car; on the lower side, that of a bus with rigid axle (left) can be compared with the case of an independent suspension (right).

At the upper side of the figure, a light duty vehicle (an off-road vehicle, for example) is represented. The drop arm, moved by the steering box, is part of an articulated quadrilateral, where the pull arms are connected or the drop arm is directly doubled. The choice between the two mechanisms is imposed by the available space and by the elasto-kinematic behavior of the suspension.

On the lower side, at the right, is shown a bus or, more generally, a vehicle in which the distance between the steering wheel and steering axle is relevant; in this case, a longitudinal pull rod, similar to that of a rigid axle (on the left), moves an additional swing arm.

Many efforts have been made in the past to minimize steering error; nevertheless, the importance of correct kinematic steering is often overestimated.

As a matter of fact, we should remember that:

- A wheel side slip angle is always present, by steering.
- Most suspensions cause an additional steer angle with roll.
- In most cases steering wheels must have a slight toe-in angle.
- Additional steering angles are caused by suspension stroke and deformation.

All these facts reduce the importance of the steering error and suggest that this issue should be considered from a broader point of view.

Steering error has, nevertheless, a significant effect on tire wear of the front wheels and on steering wheel self-alignment, influencing the driver's feeling of the steering wheel.

It is important that the reaction torque of the steering wheel increases progressively with steering angle, a feature that can be achieved with a suitable geometry.

The radius of the trajectory of the centre of mass of the vehicle is:

$$R = \sqrt{b^2 + R_1^2} = \sqrt{b^2 + l^2 \cot^2(\delta)}\,, \tag{4.9}$$

where δ is the steering angle of the equivalent two-wheeled vehicle (Fig. 4.1b), also called the *monotrace* model. Although it should be computed by averaging the cotangents of the angles of the two wheels,

$$\cot(\delta) = \frac{R_1}{l} = \frac{\cot(\delta_1) + \cot(\delta_2)}{2}, \tag{4.10}$$

it is very close to the direct average of the angles.

Consider, for example, the same vehicle of Fig. 4.2 with centre of mass at mid-wheelbase on a curve with a radius $R = 10$ m. The correct values of the steering angles are $\delta_1 = 15.090\,\text{deg}$, $\delta_2 = 13.305\,\text{deg}$ and $\delta = 14.142\,\text{deg}$. By directly averaging the steering angles of the wheels, it would follow that $\delta = 14.197$ deg, with an error of only 0.36%.

Therefore, conventionally, the steering angle δ of a mono trace model is the average of the steering angles of the wheels of the same axle.

If the radius of the trajectory is large compared to the wheelbase of the vehicle, Eq. (4.9)) reduces to:

$$R \approx l \cot(\delta) \approx \frac{l}{\delta}\,. \tag{4.11}$$

Equation (4.11) can be rewritten in the form:

$$\frac{1}{R\delta} \approx \frac{1}{l}\,. \tag{4.12}$$

The expression $1/R\delta$ has an important physical meaning: It is the ratio between the response of the vehicle, in terms of curvature $1/R$ of the trajectory, and the input which causes it. It is, therefore, a sort of transfer function for the directional control and can be referred to as *trajectory curvature gain*. In kinematic steering conditions it is equal to the reciprocal of the wheelbase.

In the case of a vehicle with independent suspension and rack and pinion steering box, the steering mechanism is quite different, as shown in Fig. 4.4.

The Jeantaud's condition could only be verified if tie rods and rack are aligned and the steering arms cross in the middle of the rear axle center line.

The complete explicit calculation of the steering error is laborious.

By using a mathematical model, we can arrive at the diagram shown in Fig. 4.6, where errors are greater compared with the previous case.

We can assume, nonetheless, that a maximum error of 5 deg is still acceptable.

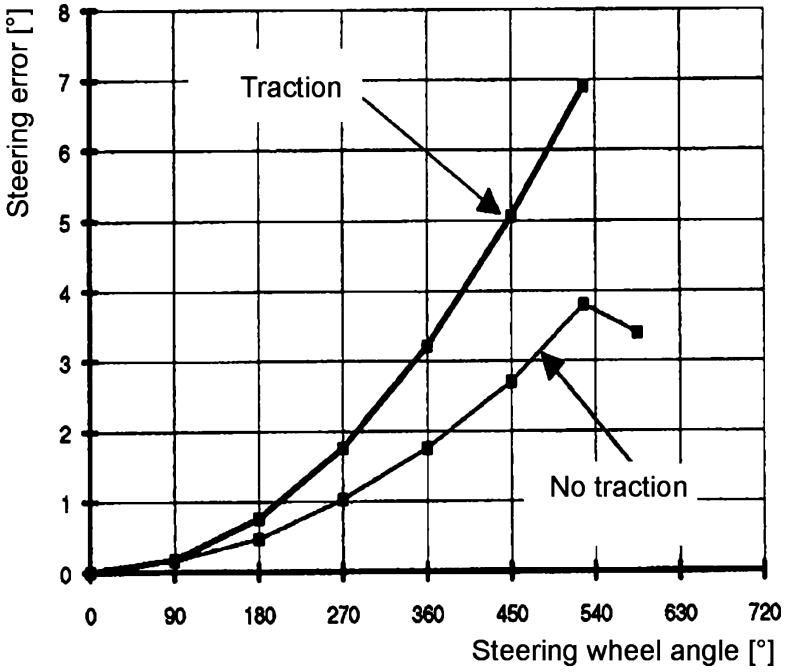


FIGURE 4.6. Steering error as a function of steering angle in a front wheel driven car, with rack and pinion steering box. Two diagrams are shown, with and without traction force.

The same figure shows a second diagram, where the steering error is calculated at the maximum traction force in this front wheel driven vehicle; this variation is the effect of the longitudinal displacement of the king-pin axis through the effect of the traction force.

The wheel steering angle must be also calculated, as a function of steering wheel angle; the result of a calculation of this kind is reported in Fig. 4.7; this result is obtained by a simple mathematical model of the articulated system representing the steering mechanism.

On the right of the diagram the right steering angle is shown for a maximum steering wheel rotation of about 360 deg, in the counterclockwise direction; on the other side, the same diagram is shown for the right wheel.

The overall transmission ratio of the steering mechanism is, therefore, about 20:1.

This value is typical for a car without power steering assistance; for a power assisted steering system this value can be reduced with the benefit of a more immediate response.

The two curves are almost linear, at least on the investigated field; they can lose this property at the rack full stroke, at about 400 deg of steering wheel rotation.

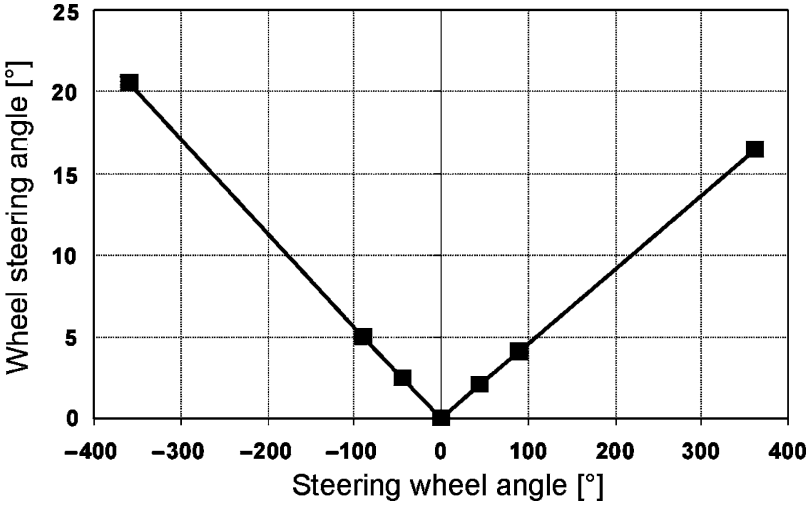


FIGURE 4.7. Diagram of wheel steering angle as a function of steering wheel angle; at the right the left wheel steering angle, at the left the right wheel steering angle.

4.3 RACK AND PINION STEERING BOX

A rack and pinion steering box is found today in almost all cars and light duty industrial vehicles; the alternative, the screw and sector steering box and its variants, is in practice reserved for heavy duty industrial vehicles or those off-road vehicles still featuring a rigid front axle.

The device shown in Fig. 4.8 transforms, through the geared couple of the pinion 3 and the rack 1, the rotary motion of the steering wheel, applied by the driver, into a linear motion of the spheric heads 2, which operate the steering mechanism.

The rack accomplishes meanwhile the task of steering wheels and track rod.

Because of the simplicity of the mechanism and the reduced friction between the teeth flanks, mechanical efficiency is usually excellent; this fact is helpful because it reduces the reaction torque on the steering wheel and gives the driver a true and accurate feeling of the existing lateral tire-road friction.

As a disadvantage, the steering transmission ratio cannot be increased beyond certain values because it is limited by tooth size. The latter is imposed by fatigue resistance of the material and by the minimum number of teeth that can be cut without interference.

The steering wheel is, therefore, always immediately responsive; this fact, generally positive, precludes the application of this mechanism to heavy vehicles. Therefore, just as power steering was once limited to luxury cars, rack and pinion was limited to small cars.

Figure 4.9 shows a cross section of this steering box through the pinion axis; this version is not power assisted. The pinion 1 is supported by the ball bearing

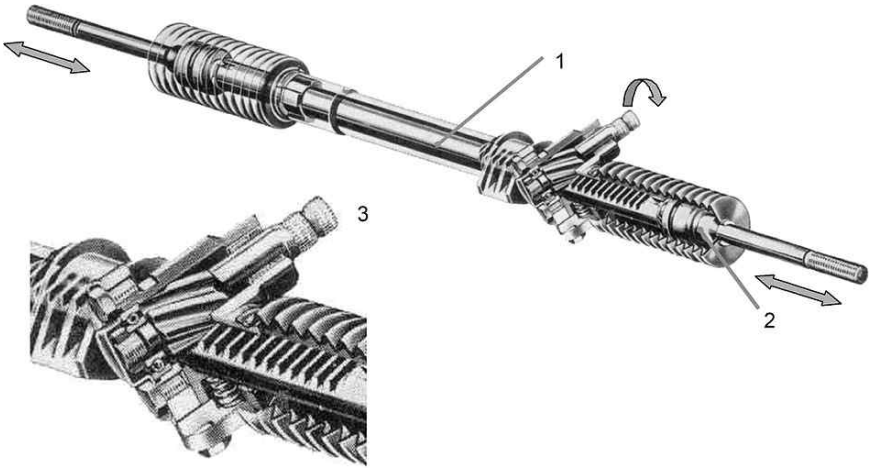


FIGURE 4.8. Rack and pinion steering box; the gearing point between rack and pinion is enlarged in the lower view.

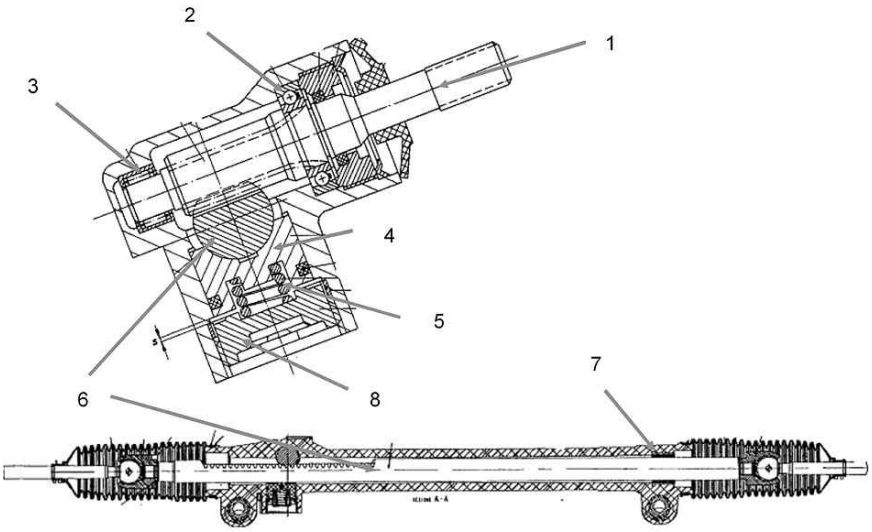


FIGURE 4.9. Enlarged detail of a rack and pinion steering box.

2 and by the needle bearing 3; the ball bearing reacts to radial and axial loads, while the needle bearing reacts to radial loads only. The rack is supported by the sliding block 4, pushed by the spring 5 that controls the pressure between rack 6 and pinion 1.

It should be noticed that the gears are of the helical type in order to allow a larger transverse contact ratio.

A sliding bush 7, made of Teflon, offers the second bearing point, at the other end of the steering rack.

The pre-loaded spring 5 overrides backlash between pinion and rack and controls the internal friction between parts. Because the sliding block 4 is mounted with a determined clearance, the spring can also absorb dynamic loads at the rack stops; the threaded ring 8, adjusted at the assembly line, determines the value of this clearance.

Rack and pinion are splash lubricated inside the tubular-shaped box; the spheric heads connecting the rack with the tie rods are set in the same oil. Two flexible rubber bellows avoid lubricant spillage and dust contamination.

The steering box is fixed to the car body or to the subframe, through the two holes shown in the same figure; these mounts feature rubber bushing, to filter noise and vibration.

The transmission ratio is determined in part by mechanism geometry, but a major role is played by the primitive radius d of the pinion.

Once the pinion is built, or a certain value for its base circle d_b is set, this simple relationship applies:

$$d = d_b / \cos\theta \ , \tag{4.13}$$

where θ represents the pressure angle of the cutting rack tool that generated the tooth wheel; by correcting the tooth profile it is possible to produce correct contact with the rack at different angles.

This feature is exploited on racks with decreasing pressure angle on each tooth, beginning with those teeth meshing with the pinion when the wheels are straight and going to those teeth that mesh with high steering angles. In this way the pitch radius increases as the steering angle increases; the result is that steering control is immediate at high speed, when steering angles are small, while steering reaction torque is reduced in parking maneuvers, when steering angles are large.

A critical point of the rack is its lateral dimension, which is determined by the necessary stroke (steering angle from stop to stop) and by a convenient distance between its bearing points; it should also be remembered that the pinion and the steering column must be positioned between the drivers feet.

The transversal bulk can sometimes affect the desired result.

To solve this problem, steering boxes with central spheric heads are designed (Fig. 4.10); in this case the box features a cut through which the two heads can emerge outside, to operate the steering mechanism; the bellow is now set inside, while the two rack ends are closed.

This configuration can be used when it is necessary to increase the length of the two tie rods to obtain a correct kinematic behavior during the suspension stroke; this happens when the steering arms point forwards instead of backwards for reasons of installation (as shown in the picture) or when the car is too narrow.

The kinematic behavior is much conditioned by rack position, but the freedom to position this component is limited.

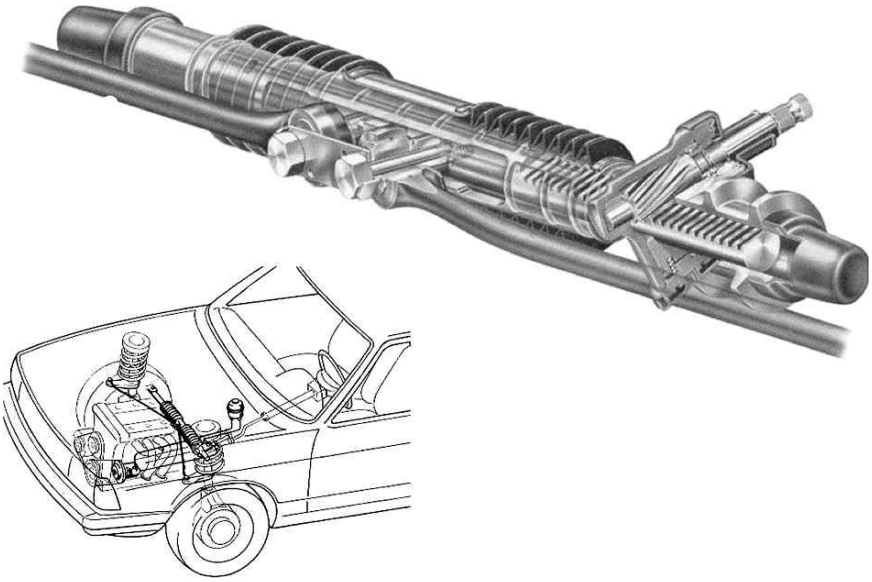


FIGURE 4.10. Rack and pinion steering box with central spheric heads, particularly useful for solving installation problems; notice also the different configuration of the steering mechanism.

The dimensions of the engine, according to its position, condition rack installation.

In cars with longitudinal engine and rear wheel drive, the rack can be installed in front of the engine, when the front axle is also in front of the engine, with steering arms pointing forwards. It can also be installed behind the wheel center, again with steering arms pointing forwards. In both cases the rack must be installed below the oil sump. In front wheel driven cars with transversal engine, the rack must be installed behind the engine, and the steering arms usually point backward.

Another constraint on rack installation is posed by the need to obtain a correct position for the steering wheel.

4.4 SCREW AND SECTOR STEERING BOX

Screw and sector steering boxes are installed, as we have explained, primarily on industrial vehicles with either rigid or independent suspensions (buses).

There are no conceptual limitations on the application of rack and pinion steering boxes to rigid axles, but we should consider that the rack should be mounted on the axle. It is difficult, in the available space, to provide for a constant velocity transmission between the pinion and the steering wheel.

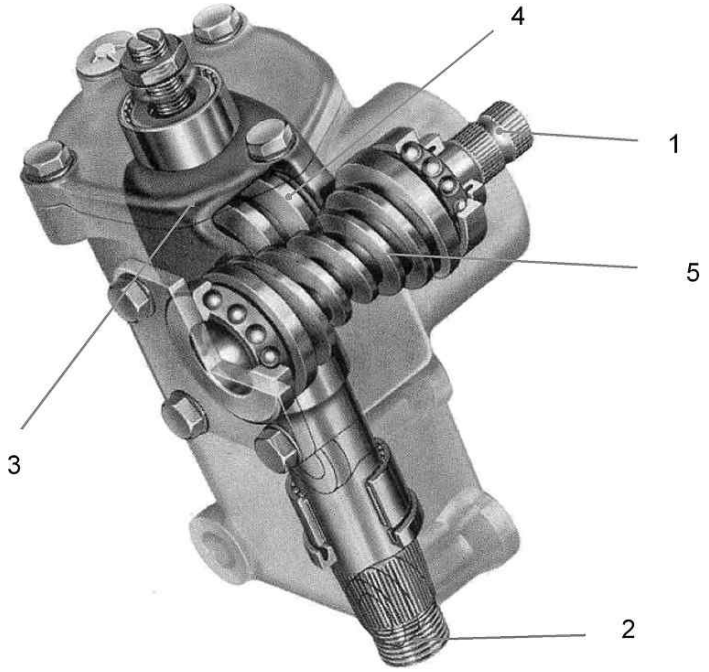


FIGURE 4.11. Screw and sector steering box. The tooth wheel sector has been made with a roller, to decrease mechanical friction. The shape of the screw copies the roller profile in order to increase the number of working threads.

Figure 3.60, in the previous chapter, shows the installation of this kind of steering box on an industrial vehicle with driver compartment on the engine. The mechanical transmission, connecting the steering wheel with the steering arm, has the configuration shown in Fig. 4.11.

This box could include a simple screw and a tooth wheel sector moving the steering arm, as often installed in the past; this solution would imply high internal friction.

Recall that this kind of transmission is all but irreversible, which means that the torque can be transmitted from the screw to the sector and not vice versa. This fact would imply lack of steering wheel self-alignment and lack of feeling for the existing friction between tires and ground.

Reversible kinematic couples, working according to the same principle and adopting a multiple threads screw, can be produced, but with much lower transmission ratios. But in this case, the advantage of reducing the steering wheel torque almost disappears.

To counteract the irreversibility, a solution has been developed where the screw 5, operated by the steering column 1, is of globoidal type and the sector shows a threaded roller 4.

The globoidal screw has a thin section in the center, to increase the number of working threads. The threaded roller, whose profile matches the screw, works

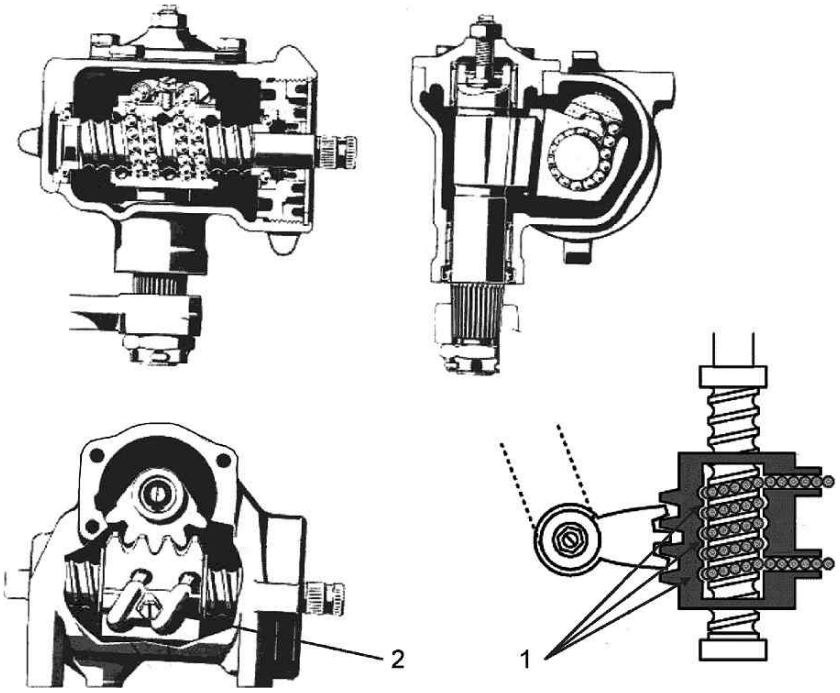


FIGURE 4.12. Screw and sector steering box with recirculating balls; this kind of screw reduces the friction to a minimum.

as a rack, capable of motion according to the screw rotation, with consistent reduction of mechanical friction.

The circumferential motion of the roller is transmitted to the spline shaft 2, through the fork 3; the shaft is connected to the steering arm.

Another way of reducing friction is shown in Fig. 4.12 with a recirculating ball screw.

In this case the globoidal screw is limited to the small portion matching the sector (it does not rotate). This screw is also threaded in the inside bore, matching with a second screw fixed to the steering column. The shape of the thread is particular, building up helical channel 1 (shown on the scheme at the lower right), inside which a number of balls can roll and move. The pipe 2 is used to close the circuit once filled with balls, to allow their recirculation.

The slip motion is again changed into rolling motion, with friction reduction; because the contact between balls and channel walls is limited, the friction is reduced to a minimum.

4.5 STEERING COLUMN

Figure 4.13 shows the assembly of a steering column for a medium size car; the miniature on the left can identify the steering column position inside the car.

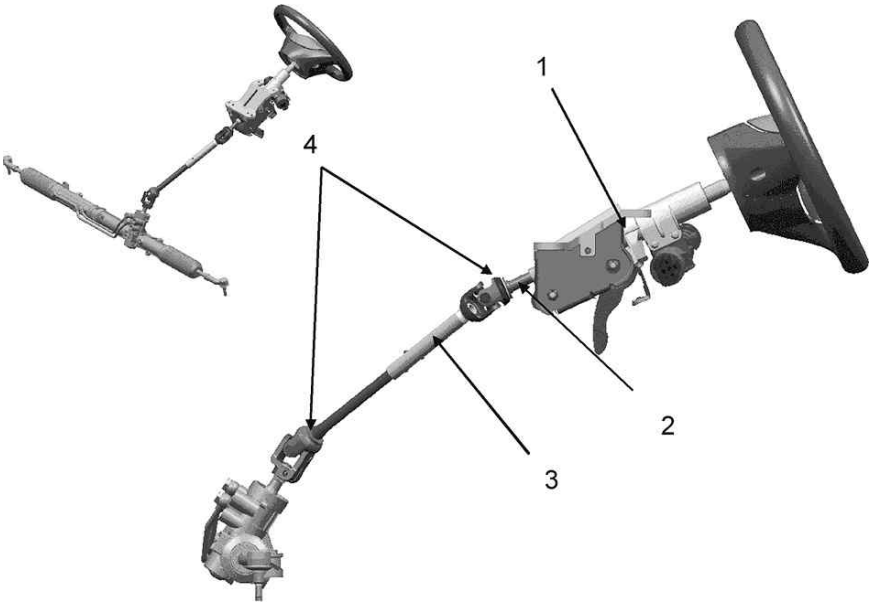


FIGURE 4.13. Steering column assembly, capable of adapting steering wheel position on the zx vertical plane. The miniature clarifies the inclined position of the middle section of the column, suitable to allow structural collapse with no intrusion of the steering wheel into the passenger compartment.

The function of this component is to transfer the torque applied to the steering wheel by the driver to the steering box. Seldom do steering box and steering wheel position in modern cars allow to have a straight steering column; for this reason the column is made up of three sections, with the first and last, respectively, connected to the steering wheel and to the steering box.

The medium section is connected to the first and the third through universal joints; the three shaft lay-out must grant a constant speed transmission. The shafts must therefore lie on the same plane and the working angles must be equal.

This kind of lay-out also has advantages for steering wheel position adjustment and passive safety.

The first section 2 of the column is fixed to the body through a structure 1, containing a pair of needle bearings or bushings. It rotates the mid-section 3 through a universal joint. The vertical position of the structure 1 can be adjusted to adapt steering wheel position to driver size.

Figure 4.14 shows a cross section of the structure 1.

It is divided into two parts: One fixed to the body (not shown in this figure), one supporting the bearing of the first section of the steering column.

This second part is connected to the first through an articulated quadrilateral, where the joints between cranks (AA' and BB') and rod A'B' allow rotation

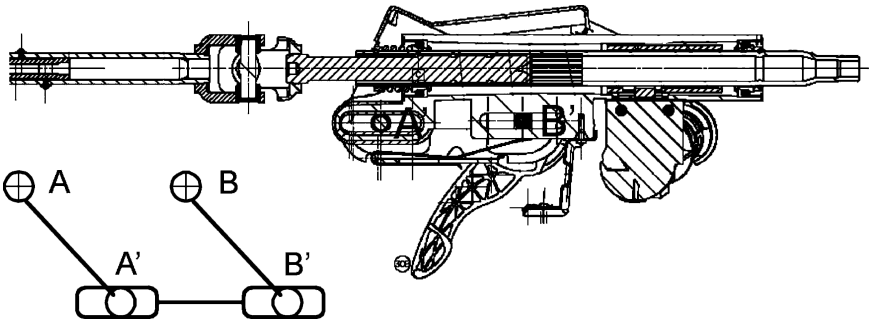


FIGURE 4.14. Cross section of the adjustable structure bearing the first section of the steering column and steering wheel: the articulated quadrilateral AA'BB' allows this adjustment.

and linear displacement, as shown in the same figure. This part can be shifted in the vertical and horizontal directions.

One of the A' and B' articulations is fastened by a screw that can fix the crank in a given position or leave the steering wheel free to be adjusted. Obviously the universal joint between the first and second section must feature a sliding spline.

Legislative regulations require the steering column to be collapsible, which means that in case of a reference crash against a barrier the steering column can reduce its length, to allow steering box displacement without steering wheel intrusion into the drivers compartment over a certain homologation limit.

This homologation requirement can be satisfied if the central section of the steering column (3 on Fig. 4.13) can change its length and if the first and third sections are not aligned.

4.6 POWER STEERING

The constant increase in passive safety requirements, comfort and convenience devices and interior room have led to a general mass increase in all cars, despite efforts devoted to applying light weight materials and design. This mass increase has affected loads between tires and ground and, therefore, the torque to be applied to the steering box to steer the front wheels.

The reduction of steering wheel torque to levels that are ergonomically acceptable has been obtained by applying power assistance to the existing mechanical systems; power steering has spread from heavy luxury cars and industrial vehicles to virtually all cars.

The most widely used power assistance is hydraulic, sometimes integrated by an electro-hydraulic device with electronic control, to adjust the effect to vehicle speed; such hydraulic systems show slight differences if they are applied to rack and pinion steering boxes or to screw and sector steering boxes.

On small and medium size cars full electrical assistance is beginning to be applied.

4.6.1 Hydraulic rack and pinion steering box

Figure 4.15 offers a general lay-out of the hydraulic power steering system for the rack and pinion box.

This system applies, as power source, the hydraulic pressure of a certain flow of oil, generated by a pump driven by the engine.

The system includes the following components:

1. Oil reservoir
2. Pump, normally blade type, driven by the accessory belt of the engine
3. High pressure tube
4. Steering box
5. Cooling serpentine
6. Low pressure return tube

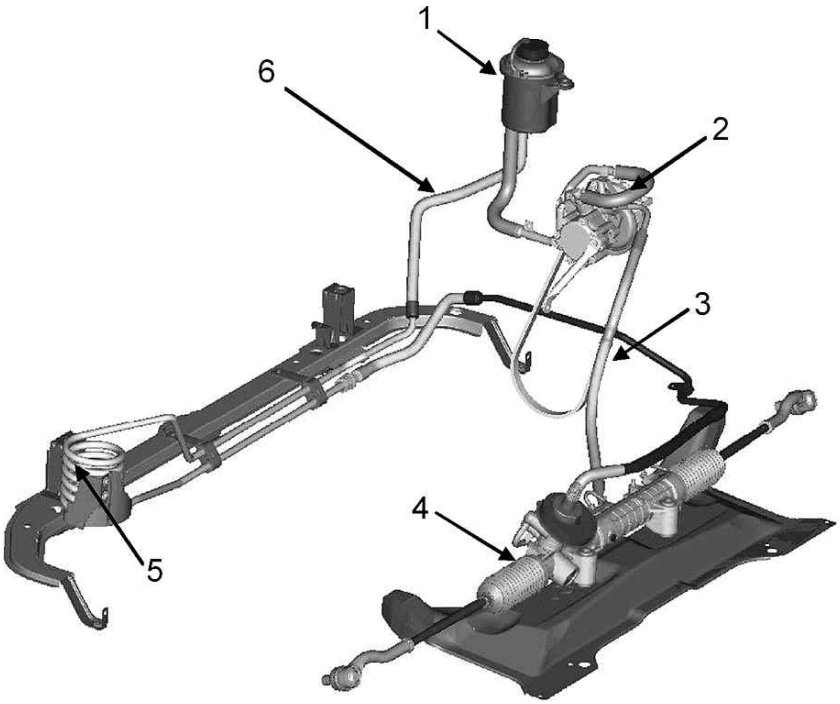


FIGURE 4.15. Hydraulic power steering system for a small front wheel driven car with rack and pinion steering box.

This scheme applies to screw and sector steering boxes as well.

The cooling serpentine is needed to waste the heat generated by the pressure control choke valves. In addition, the flow rate generated by the volumetric pump depends solely on engine speed; what is not necessary for power assistance is by passed from high to low pressure, with additional heat waste.

Inadequate cooling can allow the oil to approach the boiling point, with lack of efficiency and danger of cavitation.

Because of the choking valves and other circuit losses, this system produces a certain damping characteristic, with benefits on steering wheel return and vibrations. See also what was said about shimmy.

It appears that the relative simplicity of this system has a negative impact on fuel consumption; as a matter of fact, the fuel consumption increase is about $2\% \div 3\%$ in the average driving cycle. To limit this negative effect some systems have been developed in which the oil pump is driven by an electric motor and feeds a pressure accumulator. In this case the electric pump works only when the oil pressure in the accumulator drops below a certain threshold value, with potential reduction of hydraulic losses.

This system also has the advantage of allowing a more flexible pump installation in the engine compartments of small cars with space problems; this system had a short life and was abandoned in favour of totally electric systems.

As shown by the upper scheme in Fig. 4.16, a double effect hydraulic cylinder is integrated into the steering box, able to move the rack. M and R are the input and output nipples of the oil flow and chambers S and D on the power cylinder assist left and right steering.

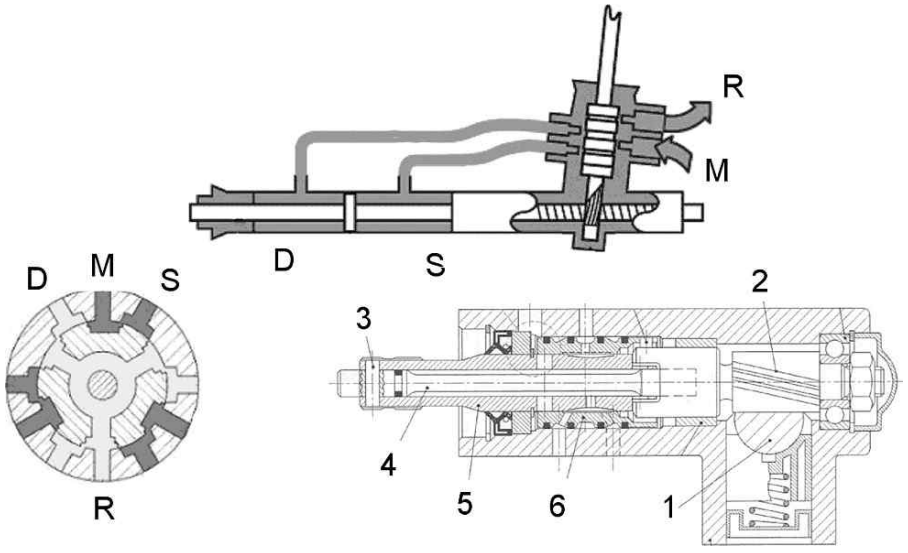


FIGURE 4.16. Schema of a hydraulically assisted rack and pinion steering box (upper side); detail of the hydraulic control valve (lower side).

The lower part of the same figure shows an enlargement of the distributor and control valve which pressurizes chambers S and D of the power cylinder; the pressure force is added to that applied to the rack by the steering wheel.

The double effect power piston can have one of the faces at the intake pressure or both at the exhaust pressure, according to the angular position of the distributor and control valve. This valve is made up of two coaxial cylinders, with relative rotation motion, featuring a number of windows, which can put the S and D chambers into communication with ports M and R, as shown in the cross section on the left.

In the position shown in the figure, the oil pressure is sent to the S chamber, while the R chamber is for exhaust. The higher the pressure, the higher the clockwise rotation angle of the inside cylinder of the valve, until a maximum opening of the window is reached; by counterclockwise rotation the pressure will be reduced until the window closes.

If the counterclockwise rotation of the cylinder is continued the same functions are performed for the other face of the piston.

The inside cylinder 5, as can be seen in the right cross section, is connected through a pin 3 to one end of a torsion spring 4, loaded by the torque applied to the pinion 1 by the steering column. In the same way the external cylinder 6 is connected to the other end of the torsion spring.

The torsional stiffness of the spring 4 must be determined in such a way that the windows are completely open at the maximum acceptable steering wheel torque. At lower torque, the spring deformation will be proportionally reduced; the steering wheel torque, even if reduced, will reproduce the aligning torque variation, thus rendering a correct feeling for the dynamic behavior of the car.

This trailing mechanism reduces the assistance torque to zero as soon as the steering wheel torque is set to zero. It should be noticed that the operation is the same when the steering wheel moves the wheels or when the wheels apply torque to the steering wheel, during steering wheel return or while driving over an asymmetric obstacle.

We should draw attention to the fact that pressure pulses or variations can generate vibrations and noise; therefore it is a good practice to apply rubber bushings to the steering box mounts and damping joints to tubes under pressure.

4.6.2 Hydraulic screw and sector steering box

Figure 4.17 shows the scheme of a hydraulic power assisted screw and sector steering box, with ball recirculation, suitable to industrial vehicles.

As can be seen, by comparing this figure with the previous, the valve scheme, here featuring two spool valves 9 and 10, moved by the fork 3, connected to the torsion spring 18, operates according to essentially the same working principle. In this steering box, the short rack 2 is also the double effect working piston, inside the cylinder 1.

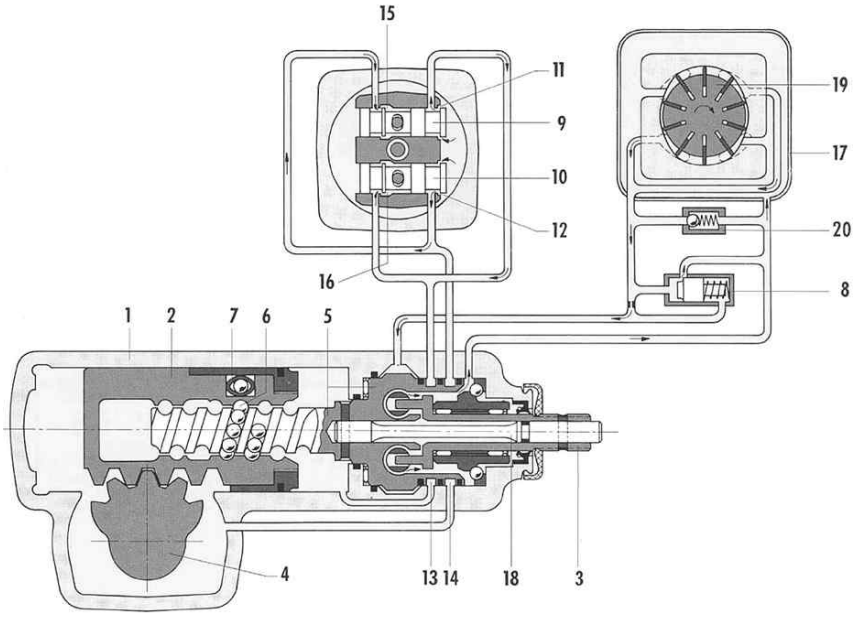


FIGURE 4.17. Cross section of a recirculating ball screw and sector steering box, with hydraulic power assistance (below) and detail of the control valve (above).

The two valve types we have shown are not necessarily limited to the steering box with which they have been presented, but can be applied to both types of box.

4.6.3 Electric power assistance

In small and medium size cars electric power steering (EPS) systems have become widespread. They are applied to rack and pinion steering boxes and operate, substantially, through an electric motor adding torque to the steering column, or to the pinion directly.

Figure 4.18 shows the assembly including a modified steering column and the detail of the electric motor with reduction gearbox.

The electric motor is regulated by an electronic controller, which again has the function of generating an assistance torque proportional to the steering torque.

The control system includes sensors capable of measuring steering torque, vehicle speed, steering wheel speed and angle.

These data are used to calculate the optimum torque to be applied by the motor; this value is a function of vehicle speed, to improve driver awareness of the smaller steering torque values. The assistance is increased at the large steering angles, typical of parking maneuvers and narrow turns, where actuation speed is more important then sensitivity.

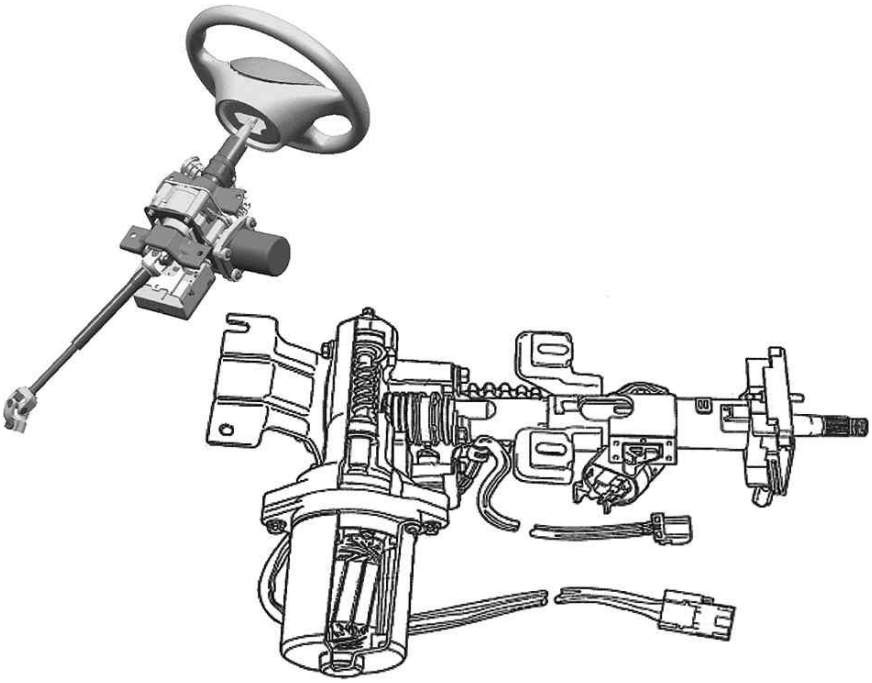


FIGURE 4.18. Electric power steering system (EPS) showing an electric motor working on the steering column. Above, a perspective view of the assembly and below a detail of the motor with reduction gearbox.

This optimum torque is applied in the correct direction, as given by steering angle acquisition.

A switch on the dashboard can provide the driver with larger or smaller assistance, as demanded by the driving environment.

The motor is a simple direct current unit with permanent magnets, driving a screw gearbox.

This system has a number of advantages when compared with a conventional hydraulic unit:

- Oil circulation is eliminated, with consequent simplification of the engine compartment lay-out.
- Fuel consumption is reduced, because torque control is not based upon wasted power.
- Active safety is increased, because assistance is also available when the engine stalls or is switched off.
- Possibility of more sophisticated torque control, as found in hydraulic systems.
- Low impact on design of mechanical components.

As a, perhaps temporary, disadvantage, there are limits on available torque, due to the working voltage, which is standardized for cars. This fact limits the application of such systems to vehicles with a total mass of about 1,500 kg; the possible adoption of a new standardized voltage of 42 V could make them available to larger cars.

With increasing production volumes new dedicated configurations have been developed, in which the electric motor works directly on the pinion or the rack.

This last application is more critical to electric motor design, working with larger torque; the electric motor works on the rack through a recirculating ball drive.

4.7 DESIGN AND TESTING

4.7.1 Outline design

A possible procedure for identifying the steering mechanism, with reference to the rack and pinion case, is outlined in the following steps; the involvement of the screw and sector can be easily extrapolated.

Let us consider Fig. 4.19.

1. The most suitable position for the rack is set according to engine and gearbox bulk and steering column lay-out.

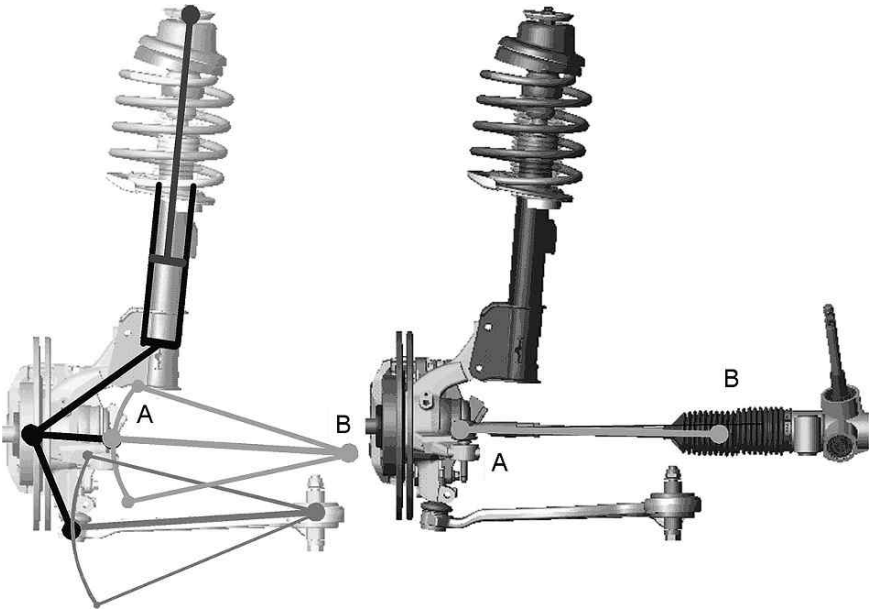


FIGURE 4.19. A useful scheme for positioning rack and steering arm.

2. From the standpoint of production, drafting may begin with the kingpin trace steering arm, attempting to satisfy the Jeanteaud condition to contain steering error. The steering arms can point backwards or forwards, but must comply with the bulk of wheel. In this way the x and y coordinates of point A are decided, while those of point B are decided according to the rack position.

3. With reference to the yz plane represented in the figure, the trace of point A can be drafted or calculated as a function of the suspension stroke, with no variation of toe angle. The path of point A can be approximated by a circle.

4. The center of this circle should be coincident with the B articulation of the steering arm, to avoid toe angle variation versus suspension stroke; if this condition is not verified the z coordinate of point A can be changed, until point B satisfies the rack installation requirements.

- By draft or calculation the angle between steering arm and steering rod must be evaluated in order to verify that the mechanism does not jam (this angle must, prudentially, occur in the interval between 20 deg and 160 deg)

- The rack length must be sufficient to allow the necessary stroke; often the two B points (right and left side of the car) are too close. In this case a rack with central spheric heads must be adopted and the procedure repeated from the beginning.

- If car understeering should be changed the z elevation of point A can be changed, until the desiderated toe angle variation is obtained.

4.7.2 Mission

The mission of the steering system includes requirements for material fatigue, crash worthiness and elasto-kinematic behaviors.

Fatigue

Designing the steering system for fatigue includes the experimentally verifying all components and the complete system. The goal of this design is to predict the life of the system under significant test conditions. The mission can be derived from that introduced for the suspension and the entire body.

In this case it is necessary to establish not only the forces acting on the wheel, but also the appropriate steering angle.

Given the importance of this system it is suggested that all parts be verified for unlimited life.

Crash

Reference is usually made to the standard test against a barrier, with increased severity, as suggested by EuroNCAP, to evaluate the intrusion into the passenger compartment and the potential danger to the driver.

Other safety targets are necessary, because the steering column also includes the driver air bag system; the driver's head must impact the inflated air bag in a pre-established way, to be protected during its travel after the impact; in other words steering column and wheel deformation must not substantially affect the deployment geometry of the air bag.

A first approximation criterion consists in evaluating whether the impact between driver's dummy and steering wheel can take place. In a positive case, injury criteria are used, limiting the dummy head acceleration after the impact. For instance, an optimum performance corresponds to a value of 72 g of acceleration, a poor performance to a value of 88 g.

Elasto-kinematic behavior

The elasto-kinematic study of the steering system includes the evaluation, as a function of the suspension stroke, of the following magnitudes at different steering angles:

- Caster angle
- King-pin angle
- Longitudinal trail
- King-pin offset

The diagrams of these magnitudes are shown in Fig. 4.20 for a front wheel driven medium size car, as a reference for a suspension of acceptable behavior.

King-pin angle variations can be small, as can those of king-pin offset; longitudinal trail, on the other hand, is primarily affected by wheel steer angle.

The design value of the caster angle will determine the design value of the longitudinal trail as well, while its variation because of the suspension stroke will affect not only longitudinal trail behavior, but also comfort and braking capacity on bumpy roads, because of the dynamic response of the sprung mass.

The design value of the king-pin angle has effects king-pin offset and steering wheel return (due to the vertical load on the steering axle), for small lateral accelerations. The king-pin angle variation will influence king-pin offset variation and the torque felt on the steering wheel during braking.

The longitudinal trail, which is very sensitive to steering angles, influences vehicle stability on straight roads and determines the steering wheel torque in curves, as well as influencing the self-aligning torque of the tire; it can, therefore, conceal the self-aligning torque decrease for high tire side slip angles, giving a different perception (through the steering wheel torque) of the friction between tire and ground.

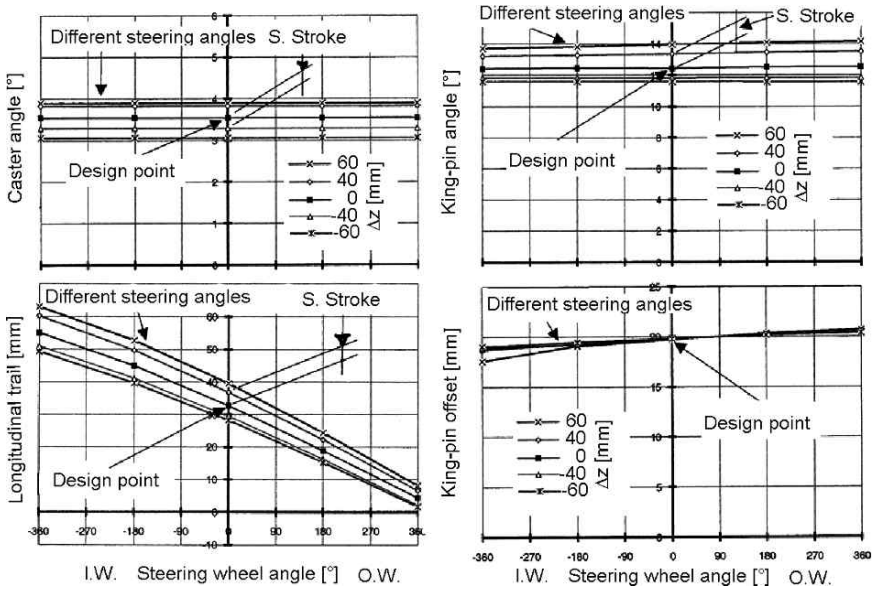


FIGURE 4.20. Diagrams of primary geometrical characteristics of the steering system as a function of suspension stroke and steering angle. I.W. (inside wheel) and O.W. (outside wheel) refer to the position of the wheel on a curved path; the outside wheel has a positive stroke.

Steering wheel torque should never be too low or negative if the cornering force can still increase when the tire side slip angle is increasing. It should be remembered that the steering wheel torque is determined by the sum of torques acting on the two wheels; what is lost on the wheel inside the path can be compensated for by an increase on the outside wheel.

King-pin offset variation influences the steering wheel torque and on vehicle stability during braking on roads with different friction coefficient on the two wheels of the same axle. As a matter of fact, in this case the torque generated on the steering wheel, different on the two wheels, is such as to direct steering to counteract the applied yaw moment.

4.7.3 Bench testing methods

The components that are usually evaluated separately are steering rods, steering box, steering column and power steering pump; with small adaptations what will be written on this last subject will also apply to electric power systems.

The steering rod is assembled to the steering box, fixed to the best test bed, and connected to a hydraulic actuator on the wheel side. These three items lie on the bench in the same position they have on the vehicle.

Fatigue tests can be performed on universal benches. Tests should be repeated on many specimens so as to obtain a sufficient statistical confidence; test results are positive if all specimens accomplish their mission without ruptures of any kind.

The steering box is usually installed on a dedicated bench, reproducing its position in the vehicle and its mechanical interfaces; the maximum mission torque is applied for a defined number of cycles (i.e. 100,000).

The mission torque is defined on a complete steering cycle from stop to stop, with the vehicle standing on good dry tarmac. During this fatigue test functional tests are repeated from time to time to verify that performance is not affected; mechanical efficiency in particular must not decrease. At the end of the test there should be no degradation or rupture.

The bench includes an electric motor with gear box and reverse gear able to apply alternate rotations in the field between ±500 deg, at a frequency of about 0.2 Hz; the torque is applied through the steering column, while the steering rods or the spheric heads are directly loaded by a pair of pneumatic actuators applying a constant reaction force; these simulate the spin friction of the tire on the road.

The steering column is tested on a fatigue bench suitable for pulsating torque. Universal joints should reproduce the working angles of the actual vehicle.

The test procedure is the same as for the steering box and, if convenient, these two tests can be performed together.

The hydraulic pump is tested on a dummy hydraulic circuit reproducing a realistic driving cycle, including only urban and curved road driving; this kind of scenario could be applied for about 50,000 km. At the end of the test there should be no rupture or loss of performance.

Similar tests can be designed for the remaining components. We will describe now some screening tests used to verify design quality and to obtain input data for use on mathematical models.

The primary synthetic parameters describing steering system performance are:

- Transmission ratio, between steering wheel and king-pin axis
- Dry and viscous friction or hysteresis
- Torque on the steering wheel as a function of the cornering force

A test system suitable for performing this kind of functional evaluation may include the following parts:

- A rotary hydraulic actuator, with angular displacement of at least 90 deg and radial load capacity of at least 30 N, with its dedicated hydraulic power generator and control system
- A universal joint without angular plays to apply forces to the steering wheel, compensating for actuator and steering wheel misalignments; high torsional stiffness and reduced inertia are also important features

- An angular displacement transducer (measurement field ±180 deg) to acquire steering wheel angle
- Four linear displacement transducers (measurement field ±20 mm) to acquire displacements of right and left wheels during steering
- Two reaction springs on the wheels able to simulate a reaction torque of about 60 Nm/deg
- Two load cells (measurement field $0 \div 1,000$ N) to acquire actual steering torque at the contact patch
- A static torsiometer (measurement field $0 \div 50$ Nm) to acquire steering wheel torque
- A signal acquisition system and dedicated data processing software

During this test, the front wheels rest on a floating plate (floating on an air bearing) with reduced friction. Screening tests are performed that impose steering angle and measure the remaining magnitudes.

Figure 4.21 shows the driver compartment of a car equipped with such a measurement system: Notice the hydraulic actuator 1, the universal joint 2, the angular displacement transducer 3 and the static torsiometer 4.

FIGURE 4.21. Driver seat of a car equipped with a measurement system to acquire steering wheel torque.

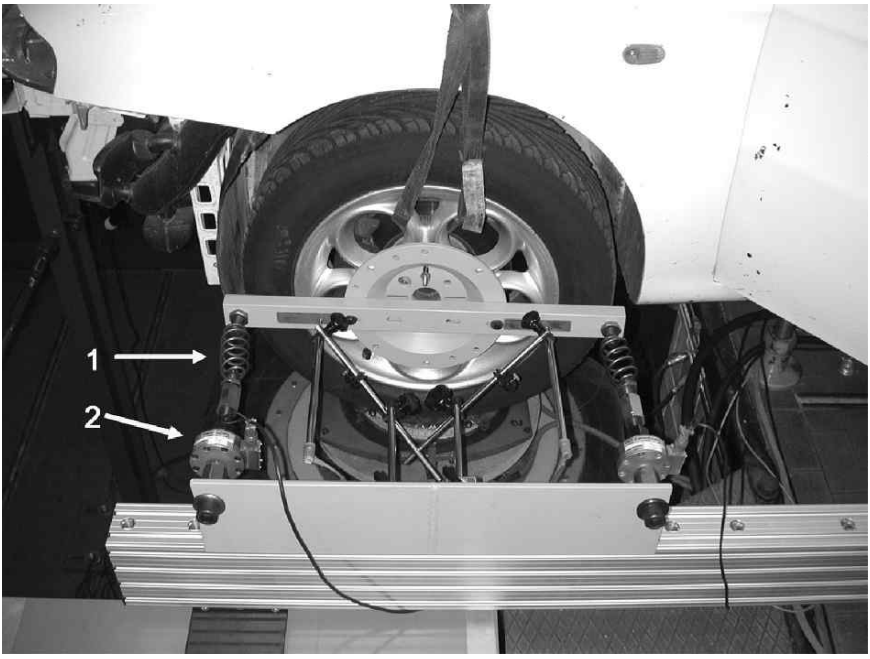


FIGURE 4.22. Part of the same measurement system of the previous picture dedicated to wheel rotation and torque acquisition. The wheels rest on floating plates.

Figure 4.22 shows the wheel side: Notice the reaction springs 1, whose effective load is measured by the load cells 2.

Steering wheel actuation is performed by sinusoidal or triangular waves of assigned frequency and amplitude, in the following test modes:

- Wheels free on floating plates
- One wheel free and one loaded
- One wheel free and one blocked

The following magnitudes are measured:

- Steering wheel angle
- Wheel steering angle
- Wheel steering torque
- Steering wheel torque

The following relevant magnitudes can be calculated by data processing:

- Steering system transmission ratio, as a function of steering wheel angle

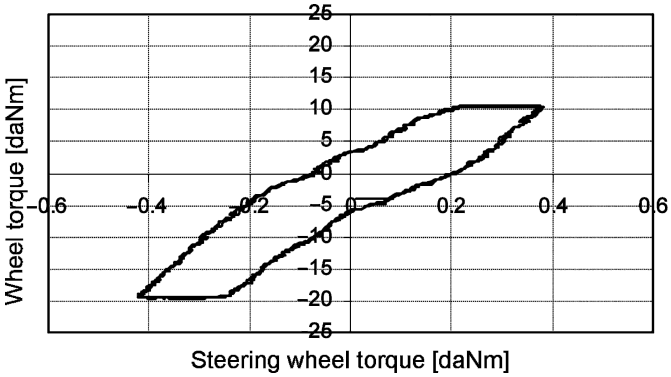


FIGURE 4.23. Steering hysteresis cycle, referred to the steering wheel.

- Steering wheel torque, as a function of angle
- Wheel steering torque as a function of steering wheel torque

This last kind of diagram is shown in Fig. 4.23. We notice a hysteresis cycle caused by mechanical friction and viscous friction, caused primarily by the power assistance system.

Dry mechanical friction is due to relative motion and is localized on spheric heads, rack with sliding block, and pinion and steering column bearings and joints. This friction is usually independent of steering angle, but is dependent on force.

The other major source of friction is the contact between tire and ground.

Viscous friction is due to oil choking and flow, localized on hydraulically assisted steering boxes; magnetic losses of electric power steering show a similar effect.

The total amount of such friction determines the hysteresis cycle; the size of this cycle shows the work performed by the driver during steering, and should be as small as possible.

5
BRAKING SYSTEM

5.1 INTRODUCTION

The braking system must accomplish three different tasks:

- To stop the vehicle completely; this function entails braking moments that are as strong as possible on the wheel.
- To control speed, when the natural deceleration of the vehicle due to mechanical friction and motion resistance is not sufficient; this function entails braking moments on the wheels that are moderate, but applied for a long time.
- To keep the vehicle stopped on a slope.

Because of the nature of these tasks the braking system is one of the safety systems of the vehicle. As a consequence, the State Authority and, later, the European Union have introduced regulations that describe design conditions and minimum operational requirements for this system.

Vehicle manufacturers and their component suppliers are, therefore, responsible for compliance of their products to regulations, including correct fabrication and system reliability for a reasonable period of time. Users, too, must play their role because many parts of this system are subject to wear and the safety functions cannot be assured without the necessary maintenance and replacement of parts. A periodic compulsory control is addressed to assessing the correct operation of this system.

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

If regulations determine minimum performance for this system, each manufacturer considers this only as a starting point, because more stringent requirements are demanded by the market and can be remarkable selling points.

Because of this, braking systems have reached in normal practice high levels of performance and reliability.

It should be remarked that the relationship between brake reliability and accident probability is not very evident; statistics on road accidents show, in fact, that less than 2% of road accidents are caused by inadequate operation of the braking system.

Within this total, 90% of accidents are estimated to be due to insufficient maintenance and 10% to dynamic instability, consequent to a braking event incompatible with transverse accelerations.

The wide application of anti-lock systems (ABS) represents an improvement for braking safety, even if accident statistics contain insufficient witness of this fact.

Studies of accidents in Germany, after the introduction of this system in 1976, showed a reduction for vehicles equipped with it; most recent data show that the presence of ABS leads users to overrate its contribution and, therefore, to expose themselves to dangers, particularly in situations such as icy roads or driving with reduced safety distance.

Another German study on taxis show no relevant difference between cars with and without ABS.

Before starting the description of braking system components, we prefer to introduce some preliminary considerations on braking system design. These will be better explained in the second volume of this book, which is dedicated to system design.

We have seen in the chapter on tires that the maximum longitudinal force exchanged with the ground depends upon many factors, the vertical load on the wheel foremost among them. This force is influenced by the nature of ground, the speed of the vehicle and the coexistence of cornering forces.

The maximum performance of this system is, therefore, conditioned by many factors outside the system itself; while the recognition of the latter is the job of drivers, who are in charge of limiting vehicle speed and controlling the distance between close vehicles, vertical loads cannot be easily understood.

These are determined by different factors, such as:

- Payload and its distribution in the vehicle
- Road slope
- Longitudinal acceleration, in particular, the same braking acceleration

To highlight this fact, let us consider a vehicle of mass m, moving on a flat road, inclined by a slope angle α.

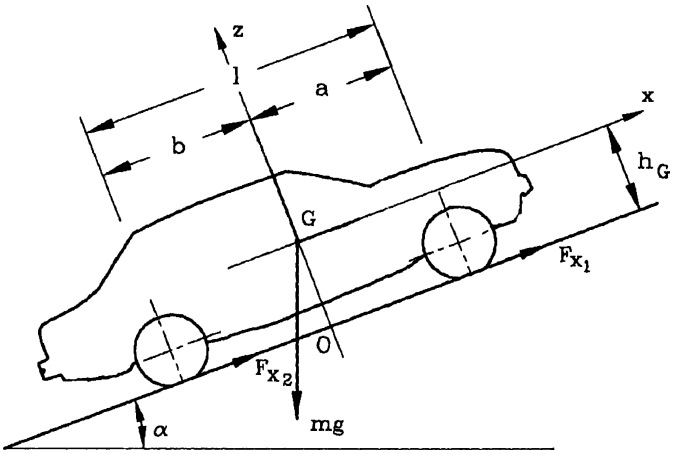


FIGURE 5.1. Scheme of forces acting on a vehicle on a flat road with slope α.

We assume that the vehicle is symmetric with reference to its median plane, while its center of gravity G has a distance a from the front axle and b from the rear axle; the center G is high h_G on the ground (see Fig. 5.1); finally:

$$l = a + b \,,$$

is the vehicle wheelbase.

We now assume that all acting forces are negligible, except for braking forces F_{x1} and F_{x2}, applied to the front and rear axle. Each of these forces represents the sum of the those acting on the wheels of the same axle, which in our assumptions are here equal.

If we indicate with F_{z1} and F_{z2} the total vertical force on each axle and with a_x the braking acceleration (it should be remarked than it is negative, because it is braking), we have:

$$\begin{cases} F_{x1} + F_{x2} - mg\sin\alpha = ma_x \\ F_{z1} + F_{z2} - mg\cos\alpha = 0 \\ F_{z1}a - F_{z2}b + mgh_G\sin\alpha = mh_Ga_x \end{cases} . \tag{5.1}$$

From this system, we obtain:

$$\begin{cases} F_{z1} = \frac{m}{l}\left(gb\cos\alpha - gh_G\sin\alpha - h_Ga_x\right) \\ F_{z2} = \frac{m}{l}\left(ga\cos\alpha + gh_G\sin\alpha + h_Ga_x\right) \end{cases} . \tag{5.2}$$

These equations indicate clearly that vertical reaction forces depend not only on the center of gravity position, but also on the vehicle braking deceleration.

In the chapter on tires we have seen that the maximum obtainable longitudinal force is proportional, through the friction coefficient μ_{xp} and the peak value of the curve $\mu_x(\sigma)$, to the vertical load on the wheel.

As a consequence, the following conclusions can be drawn:

- Maximum braking forces are determined by $m,\ b,\ a,\ h_G$ (load conditions).
- These forces are also determined by the braking deceleration a_x.
- Braking force distribution between the two axles is also determined by the obtained longitudinal acceleration.

Therefore, a braking system should be able to change its geometry to adapt braking forces to existing vertical loads. In particular the braking force on the rear axle should be progressively limited when the total braking force is increased.

This function is performed by the brake distributor valve, which will be described in the following paragraphs; the ABS system can also manage this operation with precision.

5.2 CAR BRAKES

According to regulation, the functions mentioned at the beginning of the previous paragraph are to by accomplished in a vehicle by three different systems, that cannot be directly matched to them; these are:

- *Service braking system*, able to reduce speed or stop the vehicle in normal driving conditions
- *Emergency or secondary braking system*, suitable for the same above function, but to be used in case of failure of the service brake
- *Parking braking system*, suitable for parking only, adaptable to slopes

All these systems must exert an adjustable braking force on the vehicle.

Many components are common to the first two braking systems, but some of them are specific; redundancies guarantee reliability and availability of the braking function.

5.2.1 Service and secondary systems

Figure 5.2 shows the scheme of a hydraulic braking system; it includes the service and secondary systems.

The front and rear brakes are actuated by two completely independent hydraulic circuits A and B; this separation is also present in the master pump C, actuated by the pedal and enlarged in the picture at the lower left.

This feature, imposed de facto by law, allows the two circuits to be defined, when they operate together in normal conditions, as service circuits and, when they operate separately (when one of the two has failed) as emergency circuits.

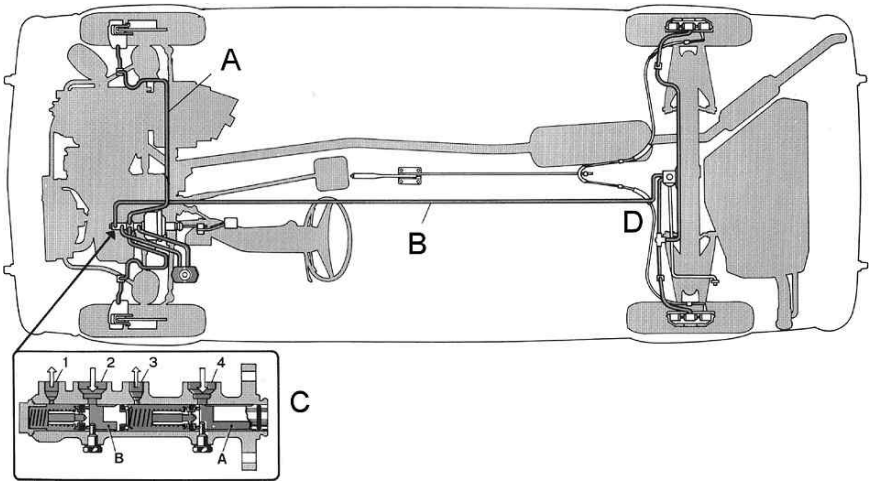


FIGURE 5.2. Pictorial scheme of a car braking system; front and rear brakes are operated by two independent hydraulic circuits A and B. C shows an enlargement of the master pump and D the distributor.

The kind of failure considered in this approach is the rupture of one of the flexible tubes connecting the brake actuators, moving with the wheels, with the rest of the circuit, fixed to the vehicle chassis structure; if, for example, the tube connecting the left front wheel should fail, at the first braking, a defined quantity of oil may be squeezed out, but the rear circuit would be still able to function.

The homologation rules for a car request that, on a perfectly paved road, this stopping distance must be obtained:

$$s \leq 0.1V + \frac{V^2}{150} \,, \tag{5.3}$$

where s is the stopping distance, measured in (m) and V is the car speed, measured in (km/h), at the onset of braking.

This condition must be met by car service system; an increased distance:

$$s \leq 0.1V + \frac{2V^2}{150} \,, \tag{5.4}$$

is allowed when one of the circuits is broken, at which point only the emergency system is functional.

Therefore, for a service circuit, an average deceleration:

$$a \geq 2.9 \text{ m/s}^2 \,,$$

is accepted, while with the emergency circuit it can be increased to:

$$a \geq 5.8 \text{ m/s}^2 \,;$$

TABLE 5.1. Braking circuits allowed by the regulations of the European Union; circuits 1 and 2 describe the still working brakes, when each of the flexible tubes, at a time, is broken.

SYMBOL	CIRCUIT 1	CIRCUIT 2
TT	Front axle	Rear axle
K	Front right and rear left wheels	Front left and rear right wheels
HT	All wheels	Front axle
LL	Front axle and rear right wheel	Front axle and rear left wheel
HH	All wheels	All wheels

these prescriptions are applicable for any load condition within those allowed by the homologation form.

The braking circuit can be organized in different ways; the schemes described in Table 5.1 are accepted.

Circuits 1 and 2 represent the possible configurations of the emergency circuit when the brake hydraulic flexible of one pipe of the wheels is out of order.

The choice among these schemes must take into account the vehicle load breakdown on the axles.

The scheme TT, for example, is suitable for a car whose center of gravity is located approximately at the middle of the wheelbase, as can happen in a front engine, rear drive two seater.

Scheme K is, instead, suitable to a front wheel driven car, where loaded with driver only, it is necessary that at least one of the front wheels can brake, to satisfy the requested performance.

The following schemes, such as HH, have the advantage of granting the emergency braking system a complete, or almost complete, braking circuit at a slightly increased cost; this fact allows the torque applied by the unbalance of braking forces to the steering wheel to be kept to a minimum.

The safety margin in *fading* conditions must factored into this choice. We will discuss fading later on; here we define fading as a loss of braking efficiency caused by a partial evaporation of the braking fluid, the result of high temperatures.

When the fluid evaporates, it increases its elasticity dramatically and loses its ability to transmit pressure by compression. This phenomenon occurs near a heat source and, therefore, near the wheel. Because of this circuits HT, LL, HH are more vulnerable when the fluid is evaporating near the front wheel.

Referring again to Fig. 5.2 we can see valve D, including the pressure distributor, which we will explain later.

In addition to the prescribed stopping distance and braking acceleration, a further condition applies to the control force needed to achieve such a stopping distance. This force, applied to the pedal by the driver's foot, must be:

$$F_p \leq 500 \text{ N} ;$$

this condition often requires the application of a power assistance device, as we will explain later.

5.2.2 Parking system

The parking braking system, operated by a hand lever or by a further pedal, is needed to keep the vehicle stopped on a slope when the driver is not in the vehicle.

The regulation specifies for this system a mechanical, non-hydraulic, connection between the driver control and the brake; tie rods and cables are allowed.[1]

The rationale for this prescription is to build up an additional emergency system, increasing system reliability by adding a backup in the rare event of a simultaneous failure of both service and emergency braking systems.

The parking braking system must be designed to keep the vehicle stopped on a slope of:

$$i \geq 18\% \ ,$$

or to reduce its speed, on a flat road, by an acceleration of:

$$a \geq 1.5 \ m/s^2 \ ,$$

after having applied on the control a force:

$$F_p \leq 500 \ N \quad ; \quad F_m \leq 400N \ ;$$

the first value applies to pedals, the second to hand levers.

The braking system must be able to generate an adjustable force that can be also maintained by driver controls; it works on the wheels of a single axle. The brakes can be the same as used for the service system.

Figure 5.3 shows a hand lever parking system: The control lever 1 moves the tie rod 17 and can be held in any position by the spring loaded pawl 27 engaging in the ratchet sector 23; the disengagement button 26 is used to unlock the pawl 27.

The tie rod 15 pulls the cable 6, through the sliding equalizer 16, dividing the force on the two brakes into two equal parts. The cable is connected to the rear brakes through the pin 8, which is connected to the two shoes.

The variant with pedal is similar; the pawl can be unlocked by a push pedal or by a button on the dashboard.

The connection with drum and disc brakes is explained in the following paragraphs.

5.2.3 Disc brakes

This brake includes a disc, rotating with the wheel, on whose two faces two brake linings made of high friction material can be pressed.

---

[1] Quite recently, the possibility of operating brakes through a bistable separate hydraulic or electric actuator has also been allowed. In this case the braking force must be maintained indefinitely with no energy.

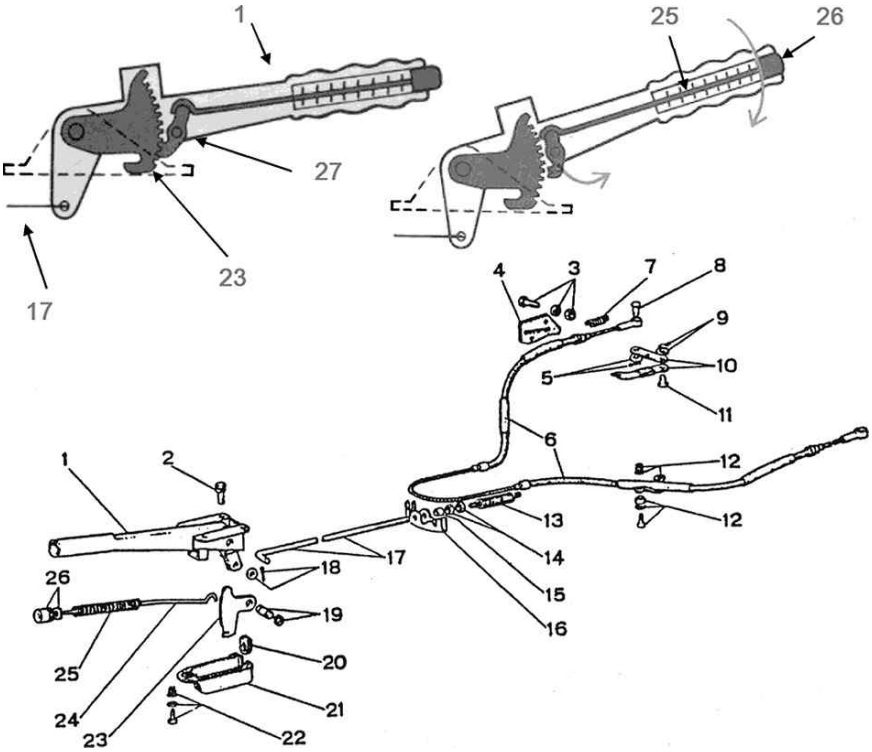


FIGURE 5.3. Scheme of a parking system with hand lever control; in the upper part of the drawing, we can see the mechanism used to maintain the braking force without need of external force.

On these linings, also called *pads*, work one or more hydraulic cylinders mounted on a suitable *caliper*.

The disc brake, like the drum brake we will describe in the following paragraph, can be mounted in two different ways, shown in Fig. 5.4, where the disc is placed:

- On the wheel hub, directly or
- On an auxiliary hub on the half shaft, at the differential box

The examples in this figure refer to a rear driving axle, but can also be applied to a front axle or a rear idling axle, by adding in this case, a shaft dedicated to the braking force.

The left figure shows the assembly of a disc brake on a guided trailing arm rear suspension. The brake caliper is fixed on the suspension strut, where the reaction to the braking force will be applied.

In the right figure, the caliper 2 is instead fixed to the differential box, or to the transmission, if differential and gearbox are integrated in a front wheel

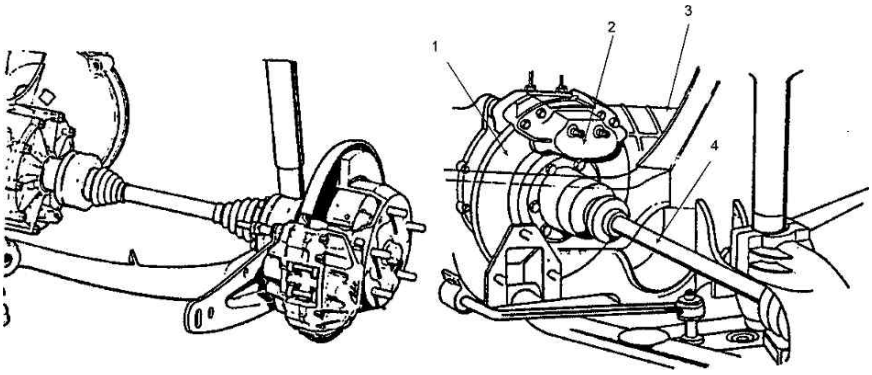


FIGURE 5.4. Examples of application of a disc brake to a rear axle; the disc can be mounted on the wheel itself (left) or attach to the differential box, working on the half shaft (right) to reduce the unsprung mass.

driven vehicle; the reaction to the braking force is applied, in this case, to the car body directly or through a suspension. The disc 1 is fit to the shaft 4 that transmits both driving and braking torque.

This kind of expensive configuration allows a reduction of sprung mass, with benefits for suspension comfort.

There are two kinds of architectures for disc brakes:

- Fixed caliper
- Sliding or floating calipers

In fixed caliper architectures, shown in Fig. 5.5, pads are pressed against the disc by two pairs of independent hydraulic cylinders inside the caliper, connected in parallel to the same pressure source; the double cylinder arrangement can be exploited to separate the service circuit from the emergency circuit.

Inside the disc are radial channels of suitable shape; these behave as a radial ventilator that activates air for brake cooling. The transversal drilling on the disc also improves the direct cooling of the disc's working surface.

In lighter cars with lower performance and therefore smaller discs, a single pair of cylinders might be sufficient. Ventilation channels and drilling could then be avoided.

Discs are made of iron alloy; in high performance cars, expensive carbo-ceramic discs are now appearing, prized for their property of surviving high operating temperatures.

The floating caliper disc brake, in Fig. 5.6 shows, instead, a single cylinder 4, working on the inside pad 5, while the outside pad 6 is pressed on the disc by the caliper body 1, able to slide on its mounts 7 along a direction perpendicular to the disc surface. This solution has the advantage of reducing cost and size, allowing the king-pin offset to be reduced to negative values.

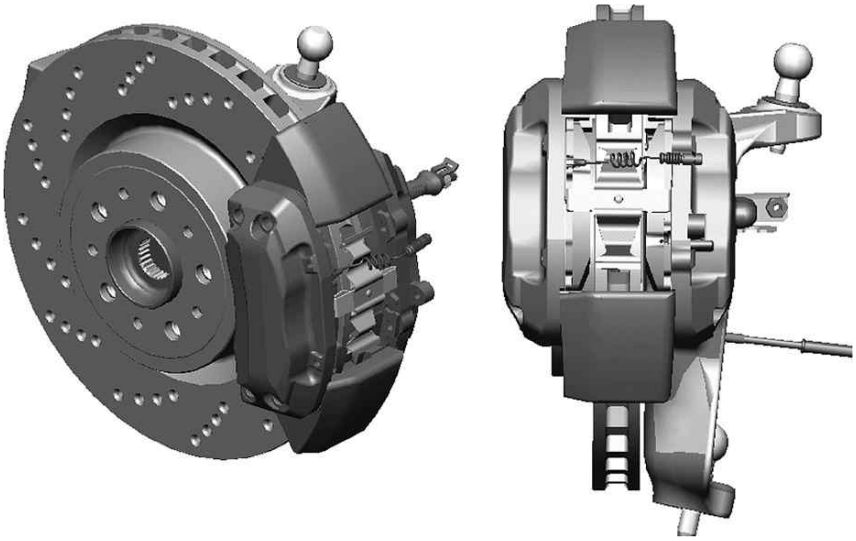


FIGURE 5.5. Ventilated disc brake with fixed caliper; the two pads are pressed against the disc by four hydraulic cylinders, two for each pad.

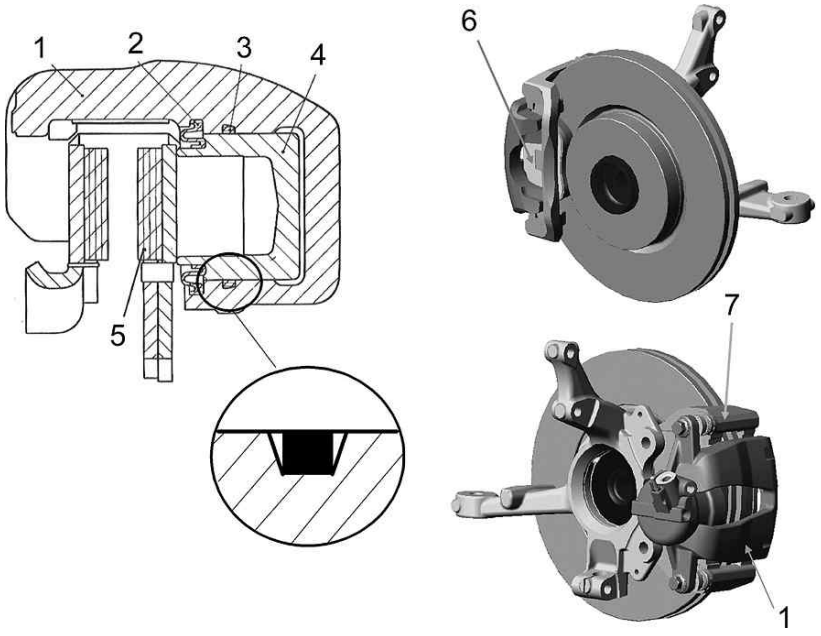


FIGURE 5.6. Floating caliper disc; the caliper is assembled so as to be able to move in a direction perpendicular to the disc surface; the hydraulic cylinder can be single. An enlarged cross section of the piston and cylinder is shown with the detail of the sealing.

Hydraulic cylinders are operated by the fluid pressure applied by the master cylinder; as soon as the pedal is released, the piston of the master cylinder returns to the initial position by means of a return spring, and the hydraulic pressure is set to zero.

The brake piston is pulled away from the disc by the elastic force, determined by the lateral deformation of the sealing ring, which is invested on each piston with a certain radial load; the shape of the groove (see the enlarged detail of the sealing ring in rest position) allows piston motion without sliding on the sealing. A similar design feature is also applied to fixed calipers.

Floating calipers have the inconvenience that cross sliding can be impaired or blocked by mud sediments or by corrosion. For this reasons protection bellows and special coatings are applied to the slide pins.

Impaired sliding of the caliper can cause different degrees of wear on internal and external pads; in some cases the caliper can be blocked by asymmetric braking, affecting vehicle path.

In disc brakes, the parking function is obtained with an additional coaxial drum brake; this solution is necessary if the force to be applied to the disc pad is too large.

This solution is shown in Fig. 5.7, at left: Inside the disc bell 1, two additional braking shoes are installed, as on drum brakes; control force is applied as will be explained in the following paragraph.

The second solution, on the right of the same figure, combines service and parking functions in a single unit; brake pads are pushed by a crank and rod mechanism 1 (the crank is made by a cam fit to the control lever 2) operated by cables. Usually this control includes an automatic backlash adjuster.

In parking brakes a difference in braking force between two wheels of the same axle is not significant for safety.

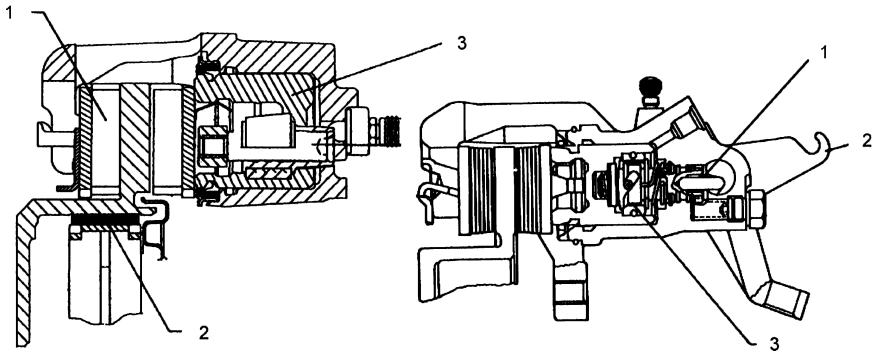


FIGURE 5.7. Different controls for parking function on disc brakes; at left, the function is obtained by an additional coaxial drum brake; at right, a lever 2 and a cam 1 act directly on the braking cylinder.

5.2.4 Drum brakes

Figure 5.8 shows a cross section of a drum brake and some design details.

This brake is made of a rotating hollow cylinder 7 (at left) fit to the wheel hub; on its inside surface work two symmetric shoes 6, on which are riveted or bonded brake linings. The shoes are pushed against the drum by the double piston cylinder 1.

At one end of the shoe one of the pistons pushes, while the other end is linked with a hinge or rests on a suitable surface. The two shoes are kept away from the drum by two springs 5.

Because of the necessary machining tolerance the drum surface and of the possible thermal deformations, the clearance between linings and drum must be well above that between lining and disc in disc brakes which, because of their simplicity, can be machined with higher accuracy and show no thermal deformations in the direction of motion of the pads.

In the same figure, on the right, a cross section of the control cylinder is shown: Two pistons 2, the oil feed drilling 3 and the return spring 6 can be seen; two pistons are assembled into the cylinder 9 and show their seals 4. The bellows 1 prevents water or dust contamination on the cylinder sliding surface.

The cylinder is connected to the oil circuit through the nipple 8; the bleeding valve 7 is used to remove air that could enter the circuit.

In modern cars drum brakes are applied to rear wheels only and supply, therefore, the mechanical parking brake control. This is performed by the push rod 3 and the crank 3, moved by cables.

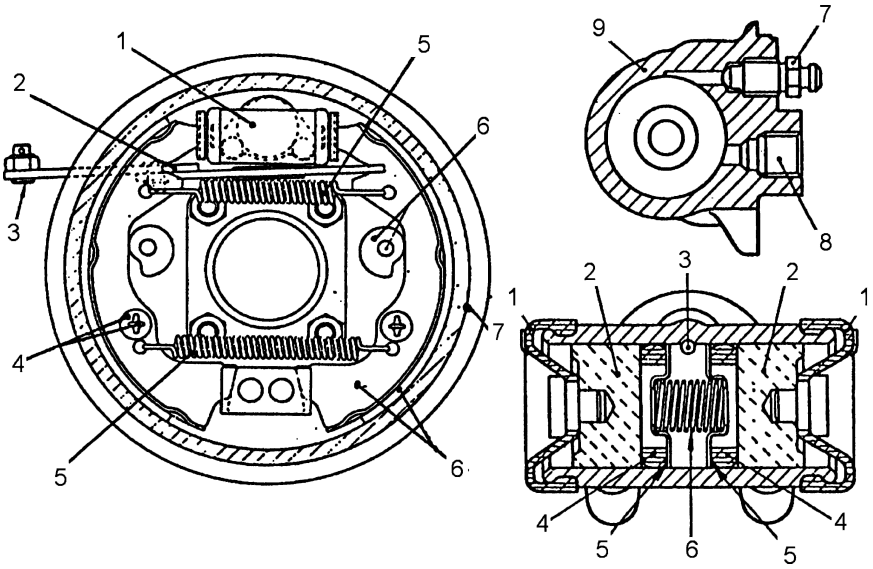


FIGURE 5.8. Section of a drum brake (at left) and of its actuation cylinder (at right).

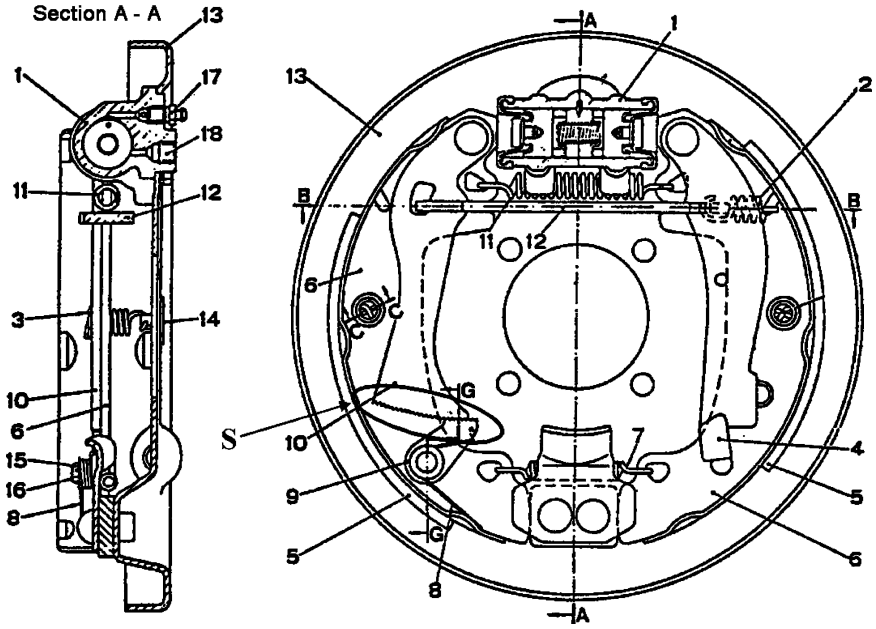


FIGURE 5.9. Section and view of a drum brake with automatic clearance adjustment. This function is accomplished by two levers 9 and 10 gearing through a tooth sector, outlined by the oval line S.

The friction between linings and drum causes them to wear. This fact is obviously present in disc brakes too, but the particular shape of the seal (see again the detail in Fig. 5.6) eliminates the effect of the additional clearance caused by wear.

In drum brakes, where the shoes are returned against a rest by a spring, this wear could cause an increase of the distance between linings and drums and a consequent increase of brake pedal stroke, which could become unacceptably large.

The rest position of the shoe must therefore be adapted to the actual wear, so as to maintain a constant clearance between shoe and drum. This adjustment can be made by manually rotating the two cams 6 by a nut and a lock nut.

In most modern brakes this adjustment is made automatically. A possible solution is shown in Fig. 5.9.

According to this system (Bendix system), one of the two shoes is made by two pieces 9 and 10 with tooth sectors gearing together, according to the detail shown within the oval line S, in this figure. The rest of the left shoe is made by the push rod 12; the clearance between rod and shoes determines the clearance between lining and drum. Until this last clearance is higher than the shoe stroke, the two pieces 9 and 10 work as a rigid system; as soon as the stroke is higher than the clearance, the upper part is retained by the push rod. This fact causes

two sectors to rotate to a different position, producing a reduction of the distance between linings and drum.

The same push rod 12 can be used for parking brake control.

5.2.5 Control system components

Pump

The pedal works on the pump piston, through the push rod on the right of Fig. 5.10.

Between this push rod and the pump the power brake is set, as we explain in the next paragraph.

The *master pump* or *tandem pump* is made by two pistons in series in the same cylinder; in this cylinder wall the openings T are connected to the oil tank, and the openings A and B to the two braking circuits. This particular pump

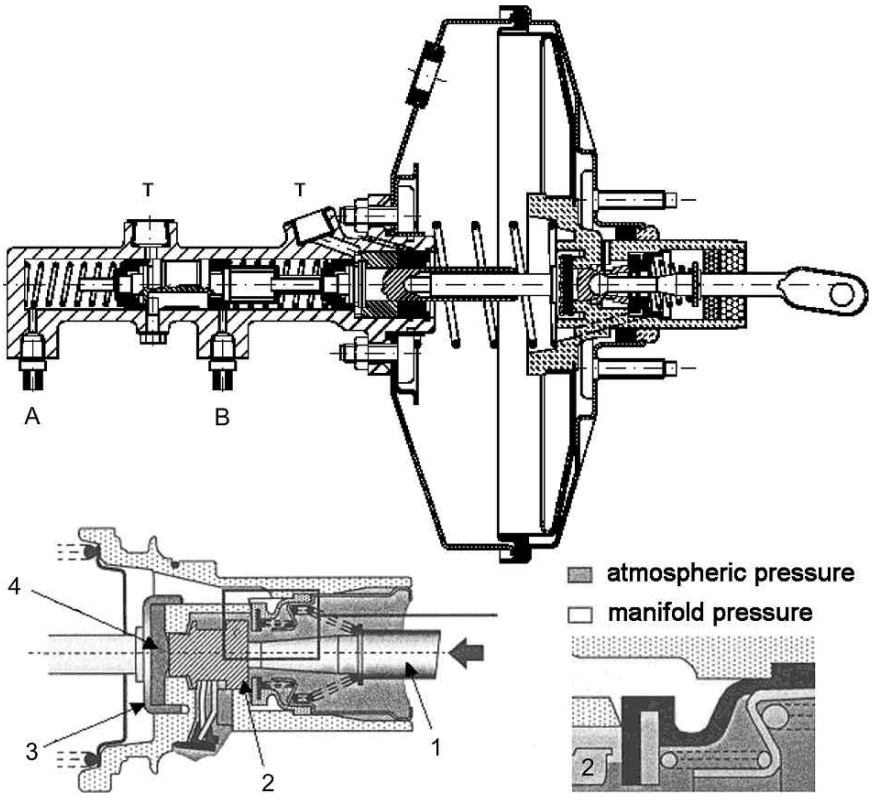


FIGURE 5.10. Cross section of the master pump and vacuum power unit. Below is shown an enlarged detail of the valve between the vacuum chamber and the ambient air, used to modulate the power assistance.

feeds two completely separated circuits (service and emergency circuits), each of which can be connected to the brakes, according to the schemes in Table 5.1.

In rest condition, the two pistons are kept to the right by the coil springs shown in the figure; the two piston chambers are connected to the tank at atmospheric pressure; in this way the additional fluid necessary to adapt to a clearance increase due to lining wear can be supplied.

As soon as the pedal is depressed the two holes T are closed and the pressure inside the circuits is increased, in proportion to the pedal force; this pressure will act on pistons working on the pads or shoes.

This kind of arrangement for the pump guarantees the operation of one circuit when the other has failed (spilling as a consequence of a pipe rupture); in fact, if one of the circuits should spill the two independent pistons will contact each other and the pressure would increase in the still working circuit. The increase in pedal stroke warns the driver about the failure.

Braking fluid

As we have seen the transmission of the pedal force to the braking surfaces is performed hydraulically.

The working fluid for this purpose must have particular features and satisfy the following specifications:

- In normal working pressure conditions the fluid must be incompressible.
- Its boiling point must exceed a certain minimum value, in order to maintain its properties after a lengthy braking.
- The fluid must have low viscosity at very low temperature, in the range of -40°C.
- It must have suitable lubrication properties for parts in relative motion (pistons, seals and cylinders).
- It must be chemically stable and non-aggressive to metal and elastomeric components.

These conditions are fulfilled by some organic oils.

These oils must be changed after a certain period of use because they are hygroscopic; water is present in the air as humidity and molecules of water can contaminate the oil in the atmospheric tank, when the level decreases as a consequence of wear.

Water in solution decreases the boiling temperature of the oil.

When brakes heat, water in solution changes to vapor; vapor bubbles decrease oil compressibility and increase pedal stroke at the same level of pressure. At critical conditions the pedal stroke is insufficient to allow the desired braking force. This lack of braking efficiency is called *fading* or *vapor-lock*.

Water absorption speed is largely dependent on climate; in humid and hot climates a percentage of 3% of the oil can be absorbed water, with a consequent 80% reduction in boiling temperature.

The United States Department of Transportation certifies fluids as DOT3, DOT4 and DOT5, defining different boiling points as a function of water content.

Distributor

Because of vertical load transfer due to vehicle deceleration, the braking force applied to the front wheels must increase as compared to the static value; for the same reason the braking force applied to the rear wheels must decrease.

Static load conditions also affect braking force distribution between the axles, because of the different position of payloads on the vehicle, with reference to the axle positions.

The function of adapting the braking force shared between axles is performed by the *brake distributor*.

The braking circuit is designed to grant the rear wheels the maximum braking force necessary, usually at full static load; the distributor is designed so as to reduce this pressure to the suitable value, corresponding to the actual static load and load transfer.

This target is achieved according to the following rules:

- When the circuit pressure is lower than a threshold value, the rear pressure is not reduced.
- When this threshold is exceeded, front pressure and rear pressure increase, according to a preset value lower than one.

This function is achieved by the valve sketched in Fig. 5.11.

The nipple 7 is connected with the pump and the front circuit, while the nipple 6 is connected with the rear circuit. In this valve a moving spool 1 responds to the rear suspension stroke, through the tip 2.

The scheme at the right of this figure shows installation on the vehicle; a suitable leverage pushes on the tip 2 when the suspension moves to compression (displacement in direction a) and vice versa in the extension direction (displacement in direction b). The vehicle suspension acts as a dynamometer measuring the axle load through the suspension stroke.

When the spool 1 is compressed in the upper direction, the valve 4 is lifted, opening the passage 3; when the spool descends, at a given position the valve 4 closes, interrupting the connection between the pump circuit and the rear brake circuit. In this condition the pressure at the nipple 6 will be reduced with reference to the pressure at nipple 7, according to the ratio between the areas on the spool 1.

The stiffness of the spring 5 determines the suspension load at which the braking pressure is reduced, while the ratio between the surfaces determines the value of this reduction.

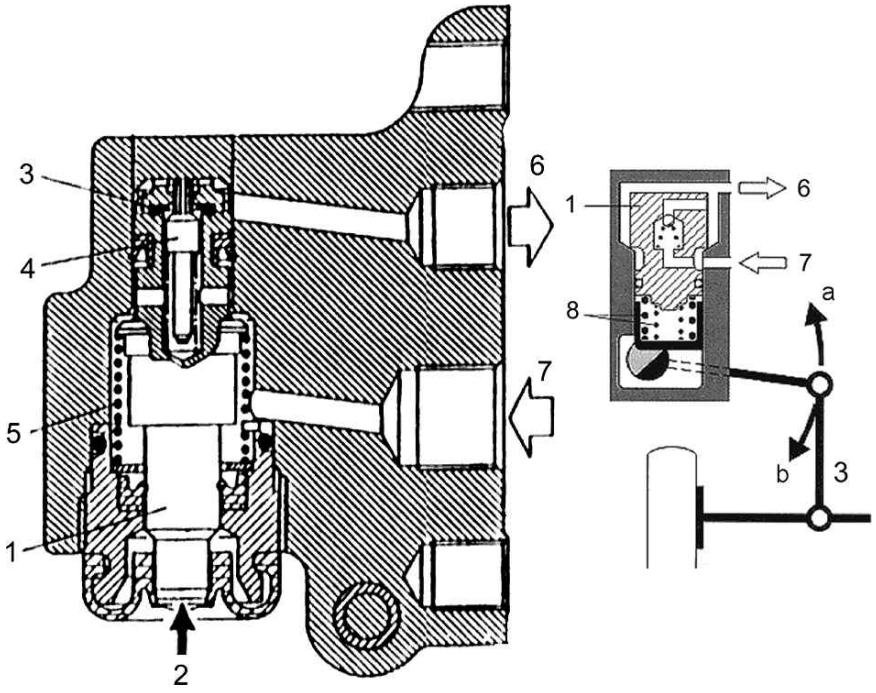


FIGURE 5.11. Cross section of a brake distributor and its installation on the rear axle.

Figure 5.12 shows an example comparing the curve of the ideal braking pressure distribution and the actual distribution made by a valve of the kind shown in Fig. 5.11. By an ideal distribution, each wheel can brake at the maximum value of the friction coefficient. The parameter for this comparison is the pressure in the front and rear circuits.

The disadvantage of a valve of this kind is that the real distribution curve is at any condition lower than the ideal, preventing the rear wheels from reaching their maximum braking capacity; on the other hand, slip is completely avoided.

A further disadvantage is that is difficult to obtain a result of this kind for any of the possible load combinations of the front and rear axle.

For reduced load variations (as on a two seater) simple pressure reducers can be used.

A third class of valves is responsive to vehicle deceleration. In this case the change of slope of the distributor curve is determined by the braking deceleration. This kind of valve takes into account the weight distribution change as well. A possible malfunction can be caused by the effect of the internal friction of the valve's mechanical accelerometer.

ABS systems also perform the function of limiting braking pressure on the rear axle, according to the actual vertical load; in this case the distributor valve is no long necessary.

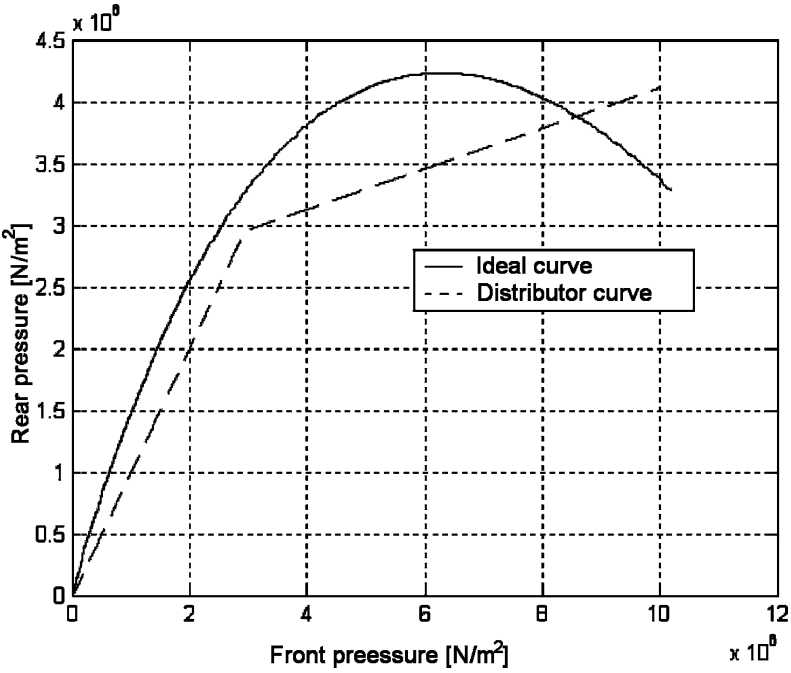


FIGURE 5.12. Comparison between the ideal distribution curve of braking pressure and the actual distributor curve.

5.2.6 Power brakes

Power brakes allow a vehicle to be braked with a reduced force on the pedal and a reduced stroke that can be contained within acceptable limits. Their advantages include braking safety and driver comfort.

To design a power brake, the following requirements must be specified.

- A power brake must be sufficiently responsive to allow the driver to modulate the braking force, even with low pedal pressure; internal friction must therefore be limited. The lower intervention point of the power system must be in the range of a pedal force of about $15 \div 20$ N.
- The effort exerted on the pedal is the feed-back of the braking system to the driver; is must be correlated to vehicle deceleration. The force on the pedal must be, in any case, proportional to braking deceleration.
- The power system response time should be lower than 0.1 s; the response time is the time necessary to reach the maximum assistance value during a sudden braking, where the pedal is depressed at a speed of about 1 m/s.
- The passage from the assisted mode to the unassisted mode at the saturation point (see the definition of this condition later on) must be gradual to allow the driver to use further force increments in emergency situations.

- The reliability of this system must be absolute; power system failures can panic the driver.
- The weight and volume must be limited to suit installation into the engine compartment.

The *tip-in load* of a power brake is the minimum load on the pedal necessary to trigger the assistance of the system during a braking event.

The *saturation load* is, instead, the value of the load at which the diagram of force on the pedal versus braking pressure changes its slope, because the assistance of the power brake has reached its maximum value. This load should never be reached in actual conditions; it must therefore be over the braking force necessary to stop the vehicle at maximum friction coefficient and maximum vehicle load.

Vacuum power brake

The power brakes is placed between brake pump and pedal and amplifies the force applied on the pedal by the driver, exploiting the difference of pressure between two chambers, one connected with ambient air and one with the intake manifold, for a throttled gasoline engine. When the vacuum manifold is not sufficient, as in diesel engines, a vacuum pump is driven by the engine.

With reference to Fig. 5.10, already used to explain the braking pump, we can note the dimension of the actuator, bound to the modest value of the difference in pressure between the manifold and the ambient air.

The actuator includes a cylinder and a piston made of steel sheet, with a suitable membrane seal; front (at left) and rear chambers are found in the actuator.

The front chamber is always in communication with the intake manifold, downstream of the throttle valve or with the vacuum pump.

Three different situations, or phases, may be identified:

- Rest position, with pedal completely released (phase 1)
- Pedal depressed (phase 2)
- Pedal released (phase 3)

Phase 1

When the pedal is released or the pedal stroke is zero (as in Fig. 5.10), the two chambers are set at the same pressure p_s. This pressure is equal to that of the vacuum source.

Because there is no pressure difference between the two faces of the power brake membrane, there will be no assistance.

In the same figure, on its lower side, is shown a detail of the shaft of the actuator's piston. On this shaft there is a valve which puts the two chambers of the power brake in communication; in this figure the valve is drafted when the pedal is depressed at zero force, or when any play is set to zero.

Phase 2

Let us now assume that a pressure is applied to the pedal. After a transient, the valve will cut the communication between the two chambers. In the detail on the left of the same figure, a plunger 1 closes the communication between the two chambers by a rubber surface.

After a short while, the rubber element 4, compressed by the braking force, will assume a given deformation, opening a passage 2 between the rear chamber and the ambient pressure.

The pressure difference between the two chambers will determine the assistance force.

The opening of the passage is a function of the deformation of the rubber element 4, sensitive to the load applied by the pedal. The passage will close as soon as the driver reaches the desired load on the pedal; the pressure in the rear chamber will therefore be proportional to the rubber element deformation and to applied load. The rubber element 4 measures the desired load.

Phase 3

When, at the end of braking, the pedal is released, communication with the outside will be closed and that between the two chambers will be reopened. Both chambers will be set at the same pressure, with no force applied to the braking pump.

In summary, the possible states of the power brake are shown in Table 5.2, as a function of the load applied to the pedal and of its derivative over time.

Figure 5.13 shows an example of the characteristics of a power brake. This curve shows the braking pressure as a function of the load on the pedal. The pressure can be calculated by the following formula:

$$p = \frac{F_s + F_p \tau}{A_p} \ , \tag{5.5}$$

where:

- F_s is the assistance supplied by the power brake;
- F_p is the force on the pedal;

TABLE 5.2. Description of the states of a vacuum power brake, as a function of the pedal force F and of its derivative over time $\frac{dF}{dt}$.

F	$\frac{dF}{dt}$	Outside duct	Inside duct	Rear pressure	Assistance
$=0$	$=0$	Open	Closed	p_s	No
>0	>0	Closed	Open	$p_s < p \leqslant p_0$	Yes
>0	$=0$	Closed	Closed	$p_s < p \leqslant p_0$	Yes
>0	>0	Open	Closed	$p_s < p < p_0$	No

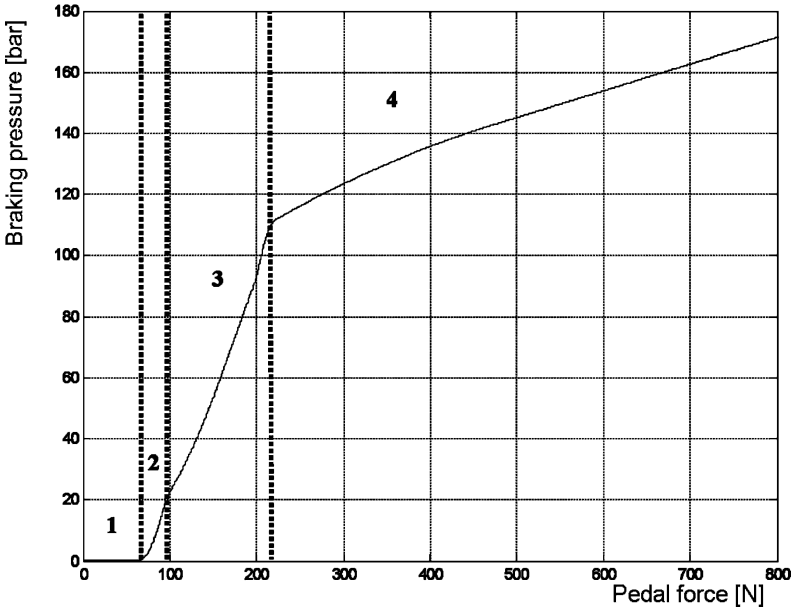


FIGURE 5.13. Example of the characteristic curve of a power brake; the braking pressure is shown as a function of the load applied on the pedal.

- τ is the leverage ratio between pedal and plunger;
- p is the braking pressure;
- A_p is the useful area of the master pump.

This characteristic curve can be measured in a bench test, where the vacuum value is kept constant, while the braking force is set to different values.

In this diagram, four zones are outlined.

- Zone 1: the braking force is insufficient to win the resistance of the springs that keep the brake in rest position. The braking pressure is thus zero. The load value, where the power brake begins to supply its contribution, is called *tip-in load.*
- Zone 2: after the tip-in load is reached there is a sudden increase of the assistance force; this phase is called *jump-in. Jump-in pressure* is the pressure value reached at the end of this phase.
- Zone 3: in this phase there is a constant amplification of the force applied by the pedal. The ratio G between the pressure and the applied load is called *power brake gain.*
- Zone 4: the power brake has reached the maximum value of pressure difference between the ambient pressure p_0 and the vacuum source p_s. The

pressure increment in this area is caused solely by an increase of the force applied by the driver to the pedal. This value is called *saturation pressure*.

Hydraulic power brake

In hydraulic power brakes the energy source is supplied by a pressurized fluid. In general, the pressure source is the same as for the power steering system and the two circuits share the same fluid. The braking servo other than the actuator is identical to what we have seen.

The higher working pressure allows a reduction in system dimensions and makes this system available to heavy cars and medium size industrial vehicles, where the vacuum pressure is insufficient.

The assistance system is made up of a simple hydraulic cylinder set in series with the master cylinder. It is fed by the power steering pump, through suitable valves.

A solenoid valve puts the power steering pump in communication with the braking circuit. When the brakes are in rest condition the pressure is available for the power steering system. During braking, priority is given to this system. A second valve modulates assistance pressure, according to the pedal force.

A pressure accumulator contains a quantity of pressure oil suitable for 2 or 3 braking, in case of the failure of the pressure source or engine stall.

For heavier vehicles, where greater storage of energy is necessary, an additional electrical pump is applied; this is used when the normal flow of oil from the steering pump is interrupted.

The power system fluid and the braking fluid are different and should not be mixed; specialized seals avoid contamination.

5.3 INDUSTRIAL VEHICLE BRAKES

European homologation regulations impose on commercial or industrial vehicles a minimum stopping distance of:

$$s \leq 0.15V + \frac{V^2}{103,5} \ , \tag{5.6}$$

where s is the stopping distance in [m] and V the initial vehicle speed in [km/h]. This formula must be applied for the service brake and, as we can see, the prescription is less stringent than for a car.

With the emergency brake, the stopping distance can rise to:

$$s \leq 0.15V + \frac{2V^2}{c} \ , \tag{5.7}$$

where c is 115 for industrial and commercial vehicles for transportation of goods and 130 for buses.

Therefore, with the emergency circuit the accepted mean deceleration is:

$$a \geq 2.2 \div 2.5 \text{ m/s}^2 ,$$

while for the service circuit it is:

$$a \geq 4.0 \text{ m/s}^2 .$$

These formulae apply for every load condition within those allowed by the homologation form.

For industrial vehicles, the so called additional retarding device (*retarder*) is also considered by regulations; this system maintains a reduced speed during long descents.

The related test procedure requires that a fully loaded vehicle be maintained for 6 km at a constant speed of 30 km/h on a downhill slope of 6%; after this test, the mean braking deceleration must be:

$$a \geq 3.3 \div 3.75 \text{ m/s}^2 ;$$

the first value applies to vehicles for transportation of goods, the second to buses.

Retarders are not part of the hydraulic or pneumatic braking systems, but are components integrated into the engine or transmission.

Those integrated into the engine are essentially devices that increase pumping losses during the intake and exhaust strokes, either by using suitable choking valves or by changing the timing of the valves.

Those integrated into the transmission are electric or hydraulic machines that use waste energy, converting mechanical work into heat that is exchanged by dedicated radiators.

We will not comment further on these devices, which are usually discussed together with engines and transmissions; the second part of this volume contains a description of hydraulic retarders.

Medium size and heavy trucks and buses have pneumatic brakes; this choice is justified by the weight of these vehicles, which cannot be braked by muscular force alone.

In such systems, the energy vector applied to actuation and assistance is compressed air, at pressures over 5 bar. A hydraulic actuated system with vacuum assistance would require, in fact, unacceptably large actuators. Nevertheless, the braking system is heavy because of the modest value of the actuation pressure, as compared with hydraulic systems; as an advantage, this fluid is unlimitedly available in the atmosphere and allows a simple design if the working fluid can be expended at each braking.

In addition, compressed air is also used for other services on the vehicle, such as automatic door opening, cabin opening for engine inspection, gearbox and clutch actuation, horns, etc.

This system is, by its nature, slower than the hydraulic alternative, but with a suitable design it is possible to deal with this drawback.

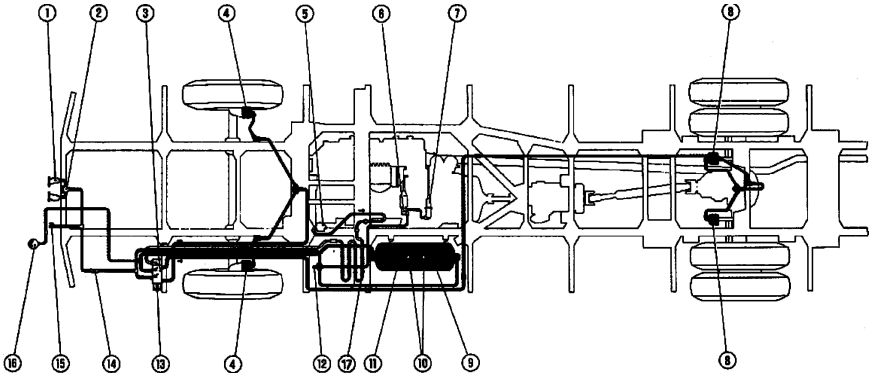


FIGURE 5.14. Scheme of a hydraulic braking system applied to a bus.

The driver's force on the pedal is used to modulate the air pressure supplied to the brake actuators. The transmission of force from pressure chambers to brake shoes is performed by levers or cams suitable for increasing forces and containing the dimensions of mechanical devices.

There are also hydro-pneumatic systems, where forces generated and controlled by compressed air are transmitted to braking pads or shoes by hydraulic pressure.

A pneumatic system is sketched in Fig. 5.14 in its simplest configuration, including a compressor 5, a control valve assembly 7, a reservoir 10, a distributor 13, a pressure gauge 16 and a number of braking actuators (4 and 8), one for each wheel.

Vehicles pulling trailers also feature a connection valve to feed the trailer braking system. In this case, regulations demand that the trailer be braked automatically with a specified performance, in case of failure of the hook or the air connection pipe.

When the engine is running, the compressor feeds the control valve assembly and the reservoir. When a preset value for the reservoir pressure is reached, the excess air is reduced to the ambient and the compressor is disengaged.

The distributor is controlled by the brake pedal and is connected to the reservoir and the brake actuators; these operate on braking pads or shoes, through a mechanic or hydraulic transmission.

When the driver depresses the brake pedal, the distributor conveys the compressed air to the braking actuator in a measure and at a pressure that are dependent on the pedal stroke; when the pedal is released, the compressed air is reduced and the brake actuator ceases its function.

In the braking circuits there are two separate sections:

- Automatic section, where the pressure is always identical with that in the air reservoir

- Controlled section, where the pressure is present only during braking

Suitable redundancies are provided to allow emergency braking in case of failures.

In the same scheme, we can also see the reservoir, divided into two sections, for front and rear brakes 9 and 11, and the serpentine tube 17 for air cooling after the compressor outlet.

Other services relying on compressed air are present: Horns 1, retarder actuator 6, an air tap to inflate tires and a pipe to feed windshield wipers 15.

The main components of this system are described in the following paragraphs.

5.3.1 Compressor

Piston volumetric compressors are used in such systems; they feature automatic disc or blade inlet and exhaust valves.

Their primary components are represented in Fig. 5.15, as the cylinder 6, the piston 5, the connecting rod 34 and the crankshaft 25; these compressors, quite similar to a small two stroke engine, apply roller bearing on all crankshaft pins.

The compressor is driven by the engine through a transmission belt. The opening of the intake valve 28 occurs when the cylinder pressure is less than

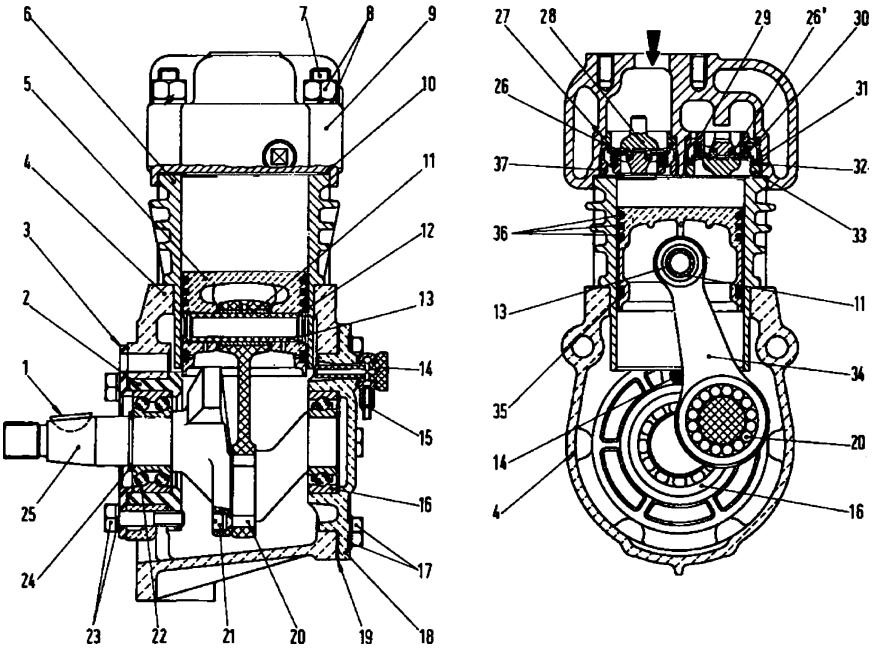


FIGURE 5.15. Main cross sections of a piston compressor for a braking system, featuring automatic valves.

the ambient pressure, while the exhaust valve 26 opens when the pressure of the cylinder is higher than the reservoir pressure.

As in thermal engines, compressor cooling can be performed by water or air, or an air cooled cylinder and water cooled head, as in the above figure; lubrication can be separated (splash type) or integrated into the engine system.

5.3.2 Control valve assembly

A control valve assembly is shown in Fig. 5.16; it includes a cleaning filter 7, a check valve 8, a pressure regulator valve 5 and a safety valve 19; it also includes a pressure gauge and a tap to inflate tires.

The filter function is to retain contaminants present in the atmosphere, while the check valve avoids reservoir download to the compressor.

The pressure regulator valve, through the exhaust valve 13, maintains the reservoir pressure below a preset maximum value.

To prevent the pressure from reaching dangerous values, a safety valve is provided: It opens when the reservoir pressure is $1 \div 2$ bar higher than the regulator valve design pressure.

When the air pressure is, instead, lower than a safety threshold, a pressure gauge switches on a warning lamp on the dashboard.

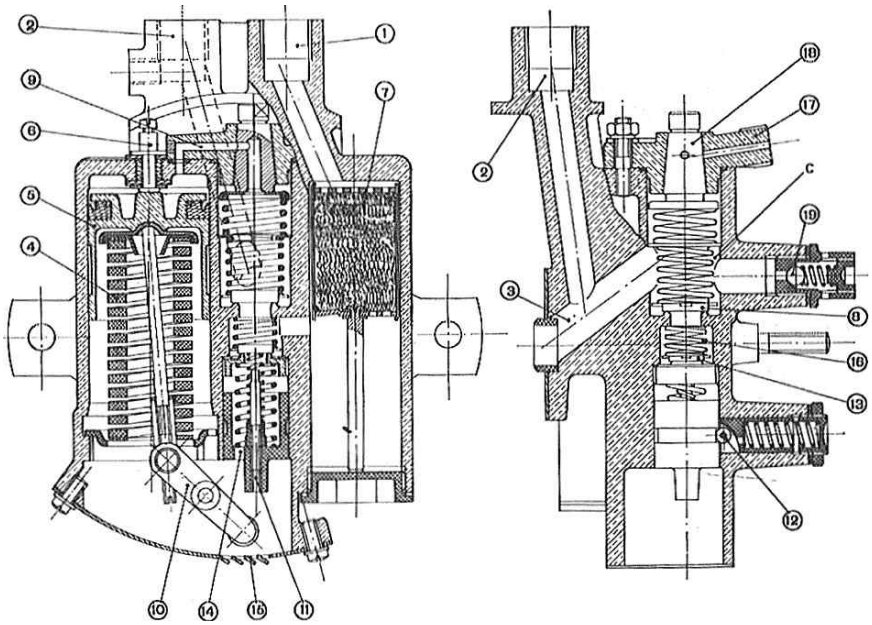


FIGURE 5.16. Cross section of a control assembly valve; this valve maintains the system pressure below a preset maximum value and warns when a minimum safety value is reached.

It is also possible to use the compressed air to inflate tires; a connection pipe line, with suitable check valves, can maintain tires at their design pressure automatically. The operation can also be done manually, when the vehicle is stopped and the engine idling.

The operation of these valves is as follows; when the engine is running the compressor feeds the reservoirs through the connection 1, the filter 7, the check valve 8, the chamber C and the duct 3; through the duct 9, the compressed air works on piston 5 of the pressure regulator.

If the air pressure is lower than a minimum safety value (usually set at 3.8 bar), the piston 5, pushed up by the spring 4 closes the switch 6, lighting the warning light.

When the pressure increases, the force on the piston consequently increases: At a certain time the piston will be pushed down and the switch will be opened. When the pressure reaches its maximum value (usually set at $5.8 \div 6$ bar), the piston 5, through the yoke 10 and the push rod 11 will open the exhaust valve 13; from this point the air coming from the compressor is downloaded to the atmosphere through the holes 14 and 15, while the connection to the reservoir is closed by the check valve 8.

When, as a consequence of braking, the reservoir pressure decreases, the piston 5 rises and the exhaust valve closes, connecting the reservoir again to the compressor.

The safety valve 19 avoids pressures higher than $7 \div 7.5$ bar.

5.3.3 Distributor

Simple distributor

Figure 5.17 shows the cross section of a simple distributor valve.

The distributor valve is made by a control piston 6, a regulation spring 7, a floating piston 8, an exhaust valve 3 and an intake valve 1.

The pressure on the pedal moves the control piston 6 and regulates the load on the regulation spring, which controls the motion of the floating piston, opening the intake and exhaust valves.

Each position of the brake pedal corresponds to a pressure in the braking circuit; the driver has feedback on the actual braking force by the reaction of the pressure in the circuit on the piston.

Let us analyze the operation of the distributor valve; by depressing the brake pedal, the piston 6 is moved and, through the spring 7, the piston 8 is also moved.

The displacement of piston 8 closes the exhaust valve 3 and opens the pre-inlet and inlet valves 2 and 1, supplying air to the braking actuators through the chamber M.

The opening of the intake valve is quick; when the pressure in the chamber M (braking actuator pressure) prevails on the spring 7, the piston 8 returns and the intake closes.

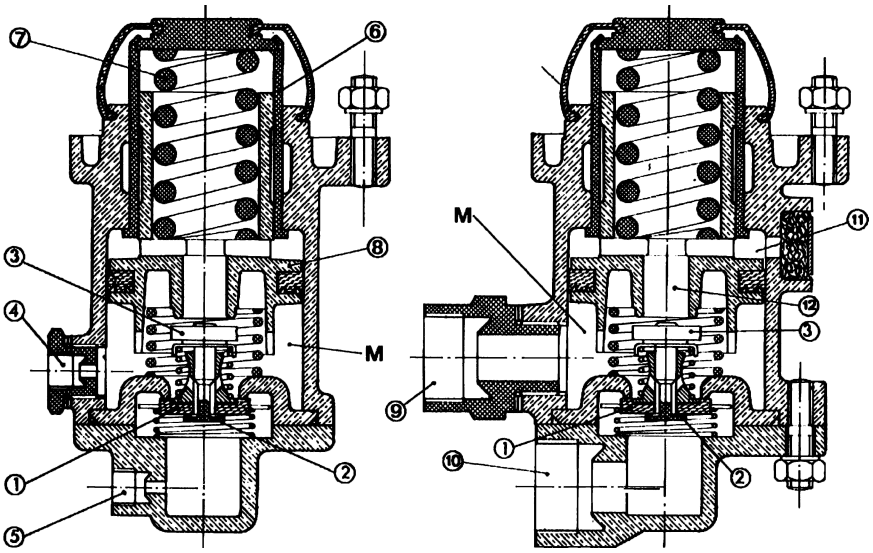


FIGURE 5.17. Cross section of a simple distributor valve, to control the pressure on brake actuators.

By reducing the force on the pedal, the force exerted by the air becomes higher than the force of the spring 7; as a consequence the piston reduces and opens the exhaust valve 3, downloading the air pressure in the chamber M through the duct 12 and the hole 11.

The pressure on the braking actuator decreases until the spring 7 is again able to close the exhaust valve.

If pressure is bleeding in the pipes or in the brake actuators, the pressure decrease in chamber M can again close the exhaust valve and open the intake valve, until the regulation pressure is again reached.

Double and triple distributors

Some braking systems feature double or triple distributors for safety reasons.

Double distributors are made up of two equal simple distributors, similar to those described in the previous paragraph; they are controlled by the same brake pedal through a yoke or a rocker arm.

The two distributors control two independent sections of the braking circuit.

If one of the sections fails, a residual braking force may remain in the other section.

Triple distributors include a third distributor; in this case, the three independent circuits can include:

- Front axle brakes
- Rear axle brakes
- Trailer brakes

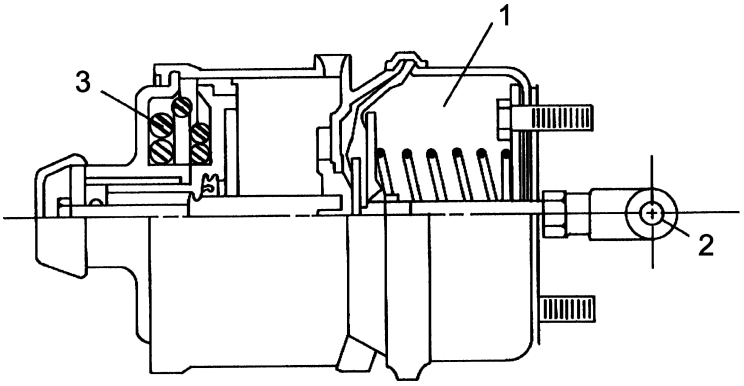


FIGURE 5.18. Combined actuator for a pneumatic brake; the left actuator is for the parking function, the right for the service function.

5.3.4 Braking actuators

Figure 5.18 shows a cross section of a combined braking actuator suitable for service and parking functions; it is made of a cylinder and a piston 2, which is connected mechanically or hydraulically to the brake shoes or pads.

During braking, the pressure air controlled by the distributor works on the actuators of each wheel; the pressure of the pads or shoes is proportional to the air pressure.

Return springs shift pistons into a rest position when the pressure is set to zero.

Because the energy source of pneumatic systems is not available for an unlimited period of time, the parking brake actuator is made by simple springs that keep the actuator normally in the brake position.

Pneumatic actuators can include a second chamber with a compressed spring, as in Fig. 5.18; in this chamber the air pressure compresses the spring and relieves the brake.

The parking brake function is therefore not influenced by the residual pressure of the reservoir.

Brakes not involved with the parking function do not feature this second actuator.

5.4 DESIGN AND TESTING

5.4.1 Braking system mechanics

Deformations

It is extremely important that brake pedal stroke does not increase over a certain limit because of obvious space limitations and because excessive pedal

displacements, during braking, are perceived by drivers as a malfunction; an excessive pedal displacement can also obstruct the application of the maximum pressure during emergency braking.

The stroke of the pedal, in the example of a disc brake, is caused by the elastic behavior of the caliper, disc, pads and pipings during braking: The oil absorption connected to those displacements contributes to an increase in the pedal stroke with reference to the theoretical value justified by pad clearance recovery.

Through experimental tests results, it has been demonstrated that caliper deformations alone cause more than 50% of the total pedal stroke.

To model this mechanism a two degrees of freedom mechanical system can be considered, where the two independent variables are the absolute displacements of the caliper x_p and of the piston x_c. This model which simply describes one half of the caliper, can be completed by including a second half.

When braking pressure is set at the atmospheric value, the system is in rest position, where the pads are set at a certain distance from the disc surface and:

$$\begin{cases} x_p = 0 \\ x_c = 0 \end{cases} . \tag{5.8}$$

When the braking process begins and pressure rises, two different phases can be identified.

Phase 1

At the beginning, only the displacement of the piston is considered, until the pads touch the disc surfaces:

$$\begin{cases} x_p = 0 \\ m_c \frac{d^2 x_c}{dt^2} + r_c \frac{dx_c}{dt} + k_c x_c - p A_c = 0 \end{cases} , \tag{5.9}$$

where:

m_c is the mass of the piston,
r_c is the damping coefficient between piston and cylinder,
k_c is the compression stiffness of the piston,
p is the circuit pressure,
A_c is the piston cross area.

This equation is applied until the following condition is true:

$$0 < x_c < c_a ,$$

where c_a is the clearance between the braking pads and the disc surfaces.

Phase II

Over this limit, the piston is considered still, while the caliper is moving, according to the equation:

$$\begin{cases} x_c = c_a \\ m_p \frac{d^2 x_p}{dt^2} + r_c \frac{dx_p}{dt} + k_c (x_p - c) + k_p x_p - p A_c = 0 \end{cases} , \tag{5.10}$$

where k_p is the caliper stiffness.

Having calculated x_p and x_c , the product:

$$A_c(x_c - x_p)$$

is the volume of oil absorbed by the deformation of half the caliper.

Therefore, the total volume of oil absorbed by caliper deformation V_p can be calculated by each half caliper contribution; we should remember that the rear brake pressure is, in general, different from that of the front brake one.

If we summarize in a single term the contributions of the three parts of the system (caliper c_p, pipes c_t and master pump c_{pd}), it is possible to evaluate the total pedal stroke c_{ped}:

$$c_{ped} = c_p + c_t + c_{pd} \ . \tag{5.11}$$

The caliper contribution c_p is calculated by dividing total oil absorption V_p by the area A of the master pump and by multiplying them by the transmission ratio of the pedal τ:

$$c_p = \frac{V_p \tau}{A} \ . \tag{5.12}$$

By determining the oil pressure in the circuit, it is easy to derive the radial and axial deformations of the tubes; pump deformation may be evaluated by experimental tests or by finite elements analysis.

Figure 5.19 shows the result of one of these tests.

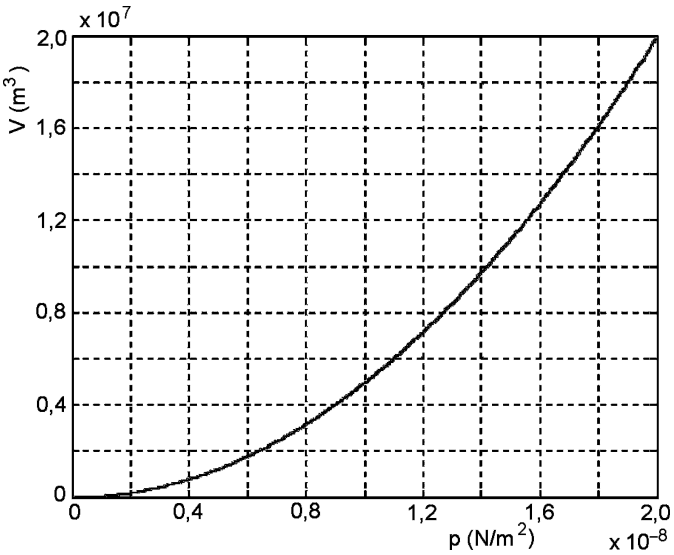


FIGURE 5.19. Diagram of the absorbed volume V in a master pump of a hydraulic braking circuit, under the pressure p.

Dynamic behavior

The dynamic response of a hydraulic system is characterized by a modest delay between input and output variables, generally lower than $0.1 \div 0.2$ s.

The dynamic behavior can be described by a static and a transient component. This last is caused by quick changes of system variables, such as circuit pressure, as a consequence of a pulse on the brake pedal. The quasi-static behavior is associated with slow variations of some characteristics, such as the friction coefficient between pads and disc caused by the wheel slowing down during braking.

Different contributions are considered in the following paragraphs.

Fluid

The flow conditions of the fluid in the hydraulic circuit, from master pump to actuator pistons, depend on viscosity, cross section and length of tubes. The parameters that determine flow speed are:

- Fluid compressibility
- Tube wall elasticity
- Flow resistance and
- Fluid inertia

Fluid viscosity contributes to increasing the time lag between force application on the pedal and a rise in braking force. It will also increase, later, the time needed to release the braking force.

In most vehicles, the tubes that feed brakes on the driver side are shorter than those on the other side, because of the position of the master pump. This fact leads to the driver side brakes being actuated more quickly than those on the other side; this difference is difficult to perceive because of the low viscosity of the braking fluid in normal conditions. But if the ambient temperature is very low, the viscosity increase can cause a yaw torque to be applied to the vehicle. The level of asymmetry of braking is influenced by the speed of application of the force to the pedal.

A viscosity increase, in addition, can increase the time needed to displace the fluid, which can induce an increase in the time needed for application of the braking force. These phenomena also affect ABS systems.

Vacuum power brake

The power brake makes a significant contribution to the response delay of braking systems. The force amplification is supplied by the motion of many different components, such as pistons, valves, springs and push rods, each of which participates in determining the dynamic behavior of the system.

In Fig. 5.20, the transient response (quick braking force application) and the quasi-steady response (slow braking force application) for a vacuum power

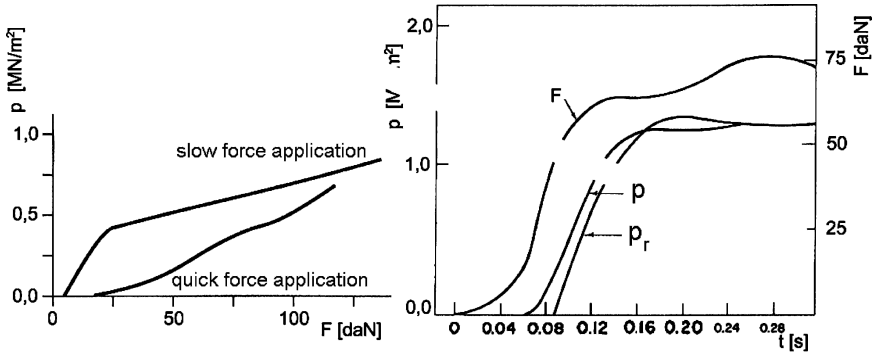


FIGURE 5.20. Left side: Quasi-steady and transient response of a vacuum power brake. Right side: Diagram of pedal force F, master cylinder pressure p and rear brake piston pressure p_r versus time.

brake are shown on the left; the input variable is the force on the pedal, the output variable the pressure in the circuit. A quick force application creates lower pressure in the circuit than that obtained by slow application, a difference that decreases only after a certain time lag.

In general, this behavior is perceived by the driver as a pedal force increase, as usually occurs when the engine is switched off.

Master cylinder

The effect of the master cylinder is generally negligible, if compared with the rest of the system, because its mass is small and its cylinder walls are very stiff.

Piping

It is possible to analyze the behavior of the tubes in the braking system with equations describing the longitudinal vibrations of the fluid in the ducts.

For small diameter tubes, the effect of viscosity is predominant.

A good correlation between calculated and experimental data can be obtained with a simple model, in which a rod, representing the fluid and set in series with a spring that represents the system elasticity, receives a force pulse at one end as input.

Brakes

The dynamic behavior of a brake (either disc or drum brake) should be analyzed by different models, considering the following behaviors:

- Thermal behavior, where the brake is considered as a device to convert kinetic energy into heat, to be discharged into the ambient air
- Mechanical behavior, considering the relationship between the braking force, as a function of the pressure and the friction between linings and brake, as a mechanical system including masses, elasticities and damping
- Friction coefficient variations

Hydraulic power brake

Braking systems with hydraulic assistance but without pressure accumulator, generally allow shorter response time than vacuum assistance, because oil is less compressible than air.

The absence of the accumulator will result in a time lag to build up the needed pressure, starting from a minimum.

In hydraulic power brakes with accumulator, the pressure is, instead, immediately accessible.

In Fig. 5.20, at the right, the dynamic behavior of a hydraulic power brake is represented; this diagram reports, as a function of time, the force F applied to the pedal, the pressure p at the master cylinder and the pressure p_r at an actuation piston of a rear wheel.

The diagrams show that pressures follow input force quite well, with a small delay due to mechanical properties.

5.4.2 Mechanical design

The mechanical design of brakes includes braking system sizing, mechanical resistance and deformation evaluation. As we have said, deformations and their consequent increase in pedal stroke should be kept as low as possible.

Thermal stresses are added to mechanical stresses: In fact, a significant part of the energy converted into heat is absorbed by the disc or the drum before being discharged into the ambient air.

In this field, finite element analysis can be fruitfully applied.

In the following paragraphs we will briefly describe the design criteria for disc and drum brakes.

Disc brakes

Modeling the system for function calculations of a disc brake is a particularly simple job.

Data for the problem are these:

f	friction coefficient between disc and linings,
R_1	pad outside diameter,
R_2	pad inside diameter,
Φ_c	piston diameter,
α	angular dimension of the pad.

The useful area of the piston is:

$$A_c = \frac{\pi}{4}\Phi_c^2 \ . \tag{5.13}$$

The useful area of the braking pad is, instead:

$$A_f = \pi(R_1^2 - R_2^2)\frac{\alpha}{360} \ , \tag{5.14}$$

where α is measured in degrees.

The braking pad exerts on the disc a pressure p_f that equals the total force F_c, acting on the pad, divided by the pad surface; if p is the pressure in the circuit, we obtain:

$$\begin{aligned} F_c &= pA_c \ , \\ p_f &= p\frac{\pi}{4}\frac{\Phi_c^2}{A_f} \ . \end{aligned} \tag{5.15}$$

In this equation p_f represents the average pressure between disc and pad and is useful for the evaluation of the stress on the pad.

If we assume that this pressure is constant at any point, the resultant force F_c will be applied at a radius:

$$R_m = \frac{2}{3}\frac{R_1^3 - R_2^3}{R_1^2 - R_2^2} \ . \tag{5.16}$$

The *brake efficiency*, usually indicated by ε, is the ratio between the braking moment and the force acting on the pads; this parameter allows an immediate performance comparison between different brakes.

If we remember that the resulting braking force is the product of the normal force by the friction coefficient, if M_f is the braking moment, we have:

$$\begin{aligned} M_f &= 2F_c f R_m \ , \\ \varepsilon &= 2f R_m \frac{\Phi_c^2}{A_f} \ . \end{aligned} \tag{5.17}$$

Drum brakes

In a disc brake the pressure exerted by the pads on the disc surface is constant in terms of direction and intensity. The formulae are consequently simple.

In drum brakes, on the other hand, the pressure distribution is not constant along the contact arch either in direction or in intensity. Consequent modelling is more complicated.

Let us consider the drum brake in Fig. 5.9 where the wheel is rotating counterclockwise; we refer to the scheme in Fig. 5.21.

The two shoes do not have the same behavior: In braking, tangent forces on the shoe are such as to push away the right shoe (which is called the *trailing shoe*) and to attract the left (which is called the *leading shoe*). The latter, when approaching the brake, rotates in the same direction as the wheel.

In particular, the leading shoe receives a braking contribution by the force of friction. In fact, if we consider the contribution of a lining element dS on the leading shoe, for the equilibrium we will arrive at:

$$F_c a = \int [p_f h dS - p_f f b dS] \ . \tag{5.18}$$

In this equation the force F_c is less than it would be for the equilibrium of normal pressures acting on the shoe alone; the braking force itself assists the pedal force.

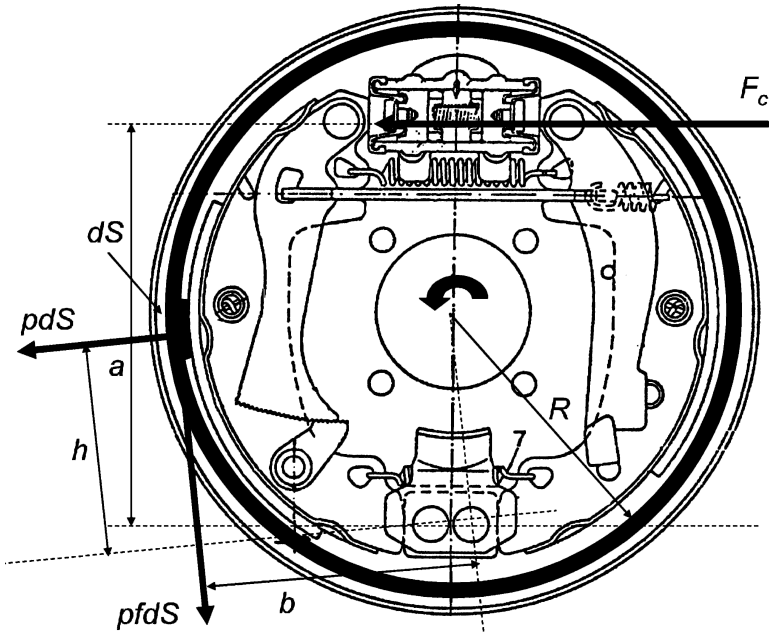


FIGURE 5.21. Scheme for evaluating the self-braking effect due to the leading and trailing shoe geometry.

For the trailing shoe, the equation will have the positive sign and the conclusion will be the opposite.

A good method for increasing the efficiency of drum brakes is to use two actuator cylinders, each of them working on a single leading shoe.

In the leading shoe the assistance is higher the higher the friction coefficient; over certain limits the brake can be self-locking when the second term in the above equations is less than or equal to zero.

To avoid this inconvenience the actual friction coefficient of the linings should be well below that of the self-locking value.

To these calculations, necessary to size the brake shoes, others are added to verify thermal and mechanical stress.

In a drum brake in particular it is usual to verify that the bending of the shoe is not excessive; good practice is to admit an additional stroke of the pedal, for shoe bending, not higher than 20% of the total.

Materials

An important issue for brake design is the friction coefficient between lining and metal. Recent studies have demonstrated that even small a variation in the alloy content of some metals, titanium, in particular, can dramatically affect the friction coefficient, as much as 20% with the same lining.

This fact has produced a common practice of applying discs and drums produced by the same casting batch on the same axle of a vehicle.

Brake linings are usually made from the following classes of component materials.

- Abrasive and solid lubricants, which impart the primary physical properties; more than one are applied, because each is active in a narrow range of temperatures. These substances are diluted with fillers with properties of mechanical and chemical resistance, but with limited abrasion or lubrication.
- Elastomers; these are added to modify physical properties, in order to increase elasticity and reduce brittleness.
- Metal powders or fibers; these improve thermal conductivity.
- Fibers; along with binders, these allow a suitable mechanical resistance to be obtained.
- Binders; these aggregate all the listed materials.

Discs and drums must show good mechanical resistance and a notable capacity for wasting heat: The most widely applied material for these purposes is grey iron.

Silicon content improves castability, producing the drawback of an increase in graphite granule size and brittleness, with an effect similar to carbon. Manganese from the metallurgical process must be limited in quantity; in combination with sulphur, it produces manganese sulfite, which impairs machinability. The maximum allowed content of manganese should be less than 1%.

For high performance cars other materials can be used, such as composites with a carbon matrix; their cost discourages application to mass production. Aluminum alloys, reinforced with silicon carbides, are also considered for their reduced weight; at best, their limited thermal resistance reduces their potential advantages in weight and size.

Because of generated heat, temperature rise on brakes is significant, especially when stopping a vehicle from high speed. As a consequence of the temperature increase, the friction coefficient will decrease, affecting brake efficiency, which will be reduced from the onset to the end of the braking.

Figure 5.22 shows a diagram of this phenomenon. It is particularly relevant in descents, where the temperature continuously increases because continuous braking does not allow sufficient time to cool down.

Efficiency can be reduced significantly and, as a consequence, pedal load can increase, as compared with the beginning of braking. This natural behavior should be partially compensated by a suitable selection of lining materials.

In heavy or high speed vehicles, self-ventilated discs are applied, as shown in Fig. 5.5.

Drum cooling can be improved by the application of outside ventilation fins, set along the circumference or parallel to the rotation axis.

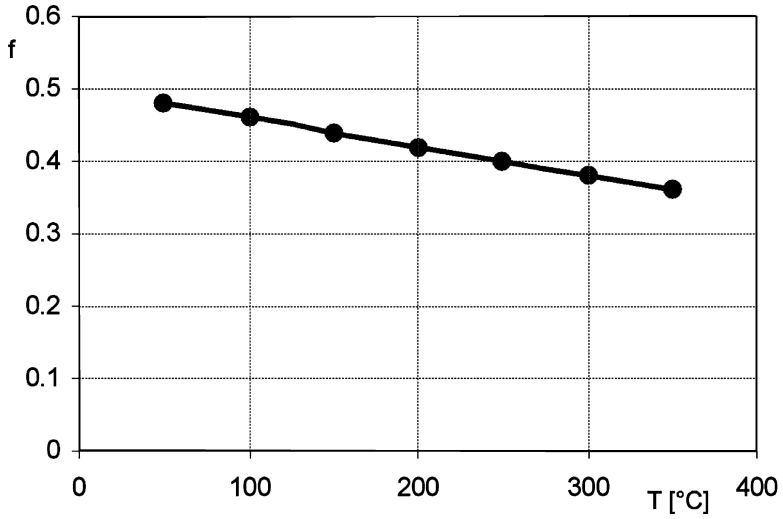


FIGURE 5.22. Diagram of the friction coefficient between pad and disc, as a function of disc temperature.

5.4.3 Thermal design

Energy and power

During braking the kinetic and potential energy of the vehicle are converted into thermal energy through the linings and brake metal surface.

If we neglect, for simplicity, the braking force exerted by driving resistance and power train motoring, the thermal energy E to be wasted during braking from the initial speed V_1 to the final speed V_2 includes the following contributions:

$$E = \frac{1}{2} m \left(V_1^2 - V_2^2\right) + \frac{1}{2} J \left(\Omega_1^2 - \Omega_2^2\right) + mg\Delta h \ , \tag{5.19}$$

where:

m	is the total vehicle mass,
J	is the moment of inertia of rotating masses referred to the wheels,
Ω_1	is the initial rotation speed of the wheels,
Ω_2	is the final rotation speed of the wheels,
Δh	is the altitude difference between the beginning and the end of the maneuver.

As a first approximation, in braking in normal conditions, tire longitudinal slip can be neglected and the rolling radius is equal to the loaded radius; therefore:

$$\Omega_1 = \frac{V_1}{R_{sc}}, \Omega_2 = \frac{V_2}{R_{sc}} \ . \tag{5.20}$$

The pressure distribution between front and rear axles changes as a function of the braking pressure and is set at a value suitable for maintaining vehicle stability during braking, thus avoiding rear wheels skid.

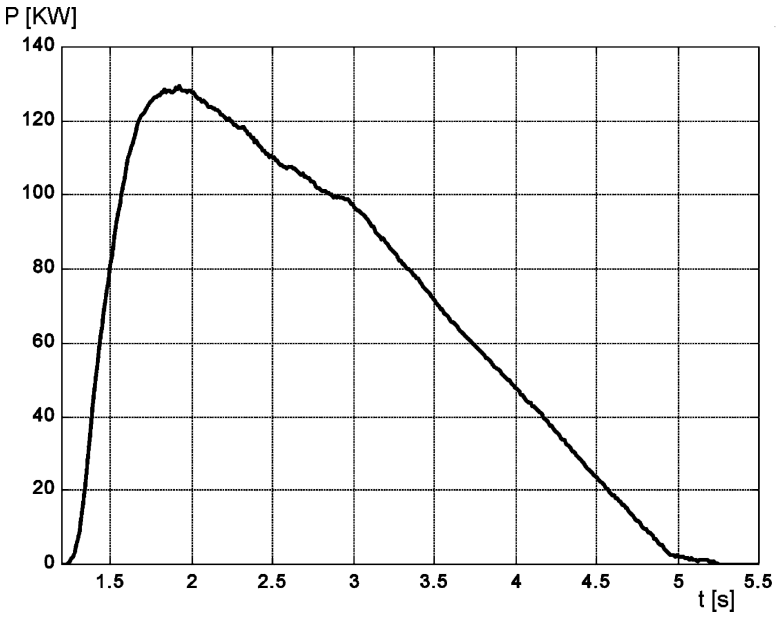


FIGURE 5.23. Diagram of the thermal power P, as a function of time t, dissipated by a front brake of a car with a mass of 1,500 kg, during a stop from 100 km/h.

If we call M_{f1} and M_{f2} the braking torques on the front and rear wheels, the equation:

$$E = 2L_{f1} + 2L_{f2} = 2\left(\int M_{f1}\Omega dt + \int M_{f2}\Omega dt\right) \tag{5.21}$$

defines the works L_{f1} and L_{f2} dissipated by brakes; this equation can be equaled to the previous.

Figure 5.23 shows, as an example, the diagram of wasted power versus time in a front brake of a car of 1,500 kg of mass, during a stop braking from 100 km/h on a ground with a high friction coefficient.

The brake thermal analysis includes not only the determination of total energy wasted by the brake, but also its breakdown between disc and pads.

This breakdown is correlated to the thermal resistance of discs (R_d) and pads (R_p), which are the elements on whose surface the heat is generated.

The heat transferred to the disc and to the liner flows through a network of different thermal resistances. In quasi-steady conditions, the phenomenon can be represented by the ratio:

$$\gamma = \frac{Q_d}{Q_p} = \frac{\sum R_p}{\sum R_d} \ . \tag{5.22}$$

where:

Q_d is the flux absorbed by the disc and
Q_p is the flux absorbed by the pad.

The previous equation implies that disc temperature and pad temperature are equal at their contact interface and that all heat generated during braking is absorbed by pad and disc only.

The ratio γ can also be written as:

$$\gamma = \frac{Q_d}{Q_p} = \sqrt{\frac{\rho_p c_p k_p}{\rho_d c_d k_d}} \tag{5.23}$$

where:

ρ is the density,
c is the specific heat,
k is the thermal conductivity.

Subscripts p and d refer, as usual, to the disc and the pad.

While we can assume that the density and conductivity of disc and pad remain constant, this assumption is not generally true for specific heat. As far as the pad is concerned, c_p is assumed to be constant, while c_d varies significantly in terms of the temperature.

The value of specific heat of disc material is shown in Fig. 5.24.

If the total generated heat, according to our assumption, is:

$$Q_d + Q_p \ ,$$

it follows that the portion of heat absorbed by the disc is:

$$\frac{\gamma}{\gamma + 1} \ ,$$

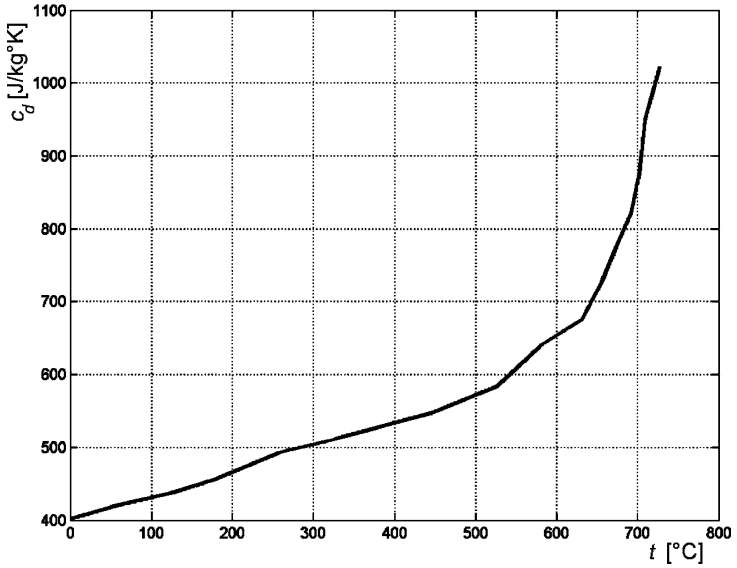


FIGURE 5.24. Diagram of specific heat of disc material, as function of temperature.

while the portion absorbed by the pad is:

$$\frac{1}{\gamma + 1} .$$

By introducing the material characteristics of disc and pad, it follows that about 80% of the friction heat during braking is absorbed by the discs.

This kind of breakdown is convenient, because too high a temperature in the pad would cause the material to deteriorate and could increase fluid temperature enough to induce local evaporation, with consequent decay of braking torque.

The heat produced during braking should be dissipated by the relative motion of the air with reference to the vehicle.

Considering the heat breakdown it appears reasonable to modify the disc geometry only, in order to improve cooling. Suitable openings on fenders can be useful in most critical cases; in the mean time the shape of the wheel disc can improve air circulation.

While modeling this phenomenon to calculate temperatures, the following heat exchanges should be taken into account:

- Convection between disc and ambient air
- Convection between disc hub and ambient air
- Convection between caliper and ambient air
- Convection between pad and ambient air
- Conduction between disc hub and wheel hub
- Conduction between disc hub and disc
- Conduction between pad and caliper
- Conduction between caliper and disc hub
- Radiation from the disc

In addition, part of the heat is absorbed by these parts as a temperature increase.

Temperature analysis

Simplified approach

During short-term braking at high speed, with high deceleration, braking time is shorter than the time needed to heat rotating parts.

In these conditions convection does not contribute to cooling and all heat energy is absorbed by the brake mechanisms and working fluid.

For drum brakes, the crossing time necessary for heat to reach the outside surface is given by:

$$t = \frac{L^2}{5a} \ , \tag{5.24}$$

where L is the drum thickness and a is the thermal diffusion coefficient, given by the formula:

$$a = \frac{k}{\rho c} \ , \tag{5.25}$$

where c is the specific heat of the drum material, k is the thermal conductivity coefficient, and ρ is the density.

The same expression for the crossing time can be used to determine the time necessary for the flux to reach the middle of the disc of a brake of this kind. In this case L would be half of the disc thickness.

For small drum brakes or ventilated discs of little thickness, the crossing time will be lower during a short braking, but in any case the heat wasted by convection will be lower than that stored in the rotor.

If we assume that braking power decreases linearly with time (constant braking deceleration), the surface temperature can be written as:

$$T(L,t) - T_i = \sqrt{\frac{5}{4}\frac{q_0}{kat}}(1 - \frac{2t}{3t_s}) \ , \tag{5.26}$$

where T_i is the initial brake temperature, q_0 is the thermal flux to the drum or the disc, immediately after the start of braking, and t_s is the vehicle stopping time.

We should notice that q_0 also represents the braking power per unit of surface, absorbed by the drum or the disc.

If we solve the previous equation with reference to time, the maximum surface temperature will be reached at:

$$t = t_s/2 \ . \tag{5.27}$$

Therefore, the maximum temperature in this kind of short braking will be:

$$T_{\max,L} = \sqrt{\frac{5}{18}\frac{q_0 t_s}{\rho c k}} \ . \tag{5.28}$$

From the previous formula we note that, for a given heat flux, maximum temperature decreases as density, specific heat and thermal conductivity increase.

Table 5.3 shows the average properties of materials used on pads, shoes and rotors of disc and drum brakes.

Second approximation approach

A better approximation can be introduced by assuming the following:

- The air temperature is constant and equal to T_∞.

TABLE 5.3. Typical values of the thermal properties of material used for drum and disc brakes.

	Pads	Shoes	Rotors
ρ [kg/m^3]	2,030	2,600	7,230
c [Nm/kg°K]	1,260	1,470	419
k [Nm/mh°K]	4,170	4,360	174
a [m^2/h]	0.00163	0.0011	0.0576

- The wheel hub is considered as a sink at temperature T_m.
- Convection coefficients are a function of vehicle speed alone.
- Conduction coefficients are assumed to be constant.
- Pad and disc specific heats are a function of the temperature alone.
- The surface between pad and disc is the sole heat source.
- Radiation invests ambient air only.
- The radiated heat in included into the convection term.

We should thus concentrate our attention on the disc, because the most relevant part of the generated heat flows through it.

To calculate convection the disc can be modeled with concentrated parameters. The disc is a single mass whose temperature is everywhere constant; disc temperature is solely a function of time and not of space.

The convection coefficient is derived from tests in which discs are cooled by air. In these tests the discs are mounted on complete cars to take into account the covering effect of the car body.

Although the disc exchanges heat by conduction with nearby parts, such as the wheel hub, most of the heat is wasted by convection.

The thermal balance is performed by the following equation:

$$\rho c V_{eff} \frac{dT}{dt} = -S(H_a + H_i)(T - T_\infty) \ , \tag{5.29}$$

where:

S is the disc surface;
V_{eff} is the disc volume, relevant for cooling;
H_a is the convection coefficient between air and disc;
H_i is the radiation coefficient.

We recall that:

$$S = 2\pi(R_e^2 - R_i^2) + 2\pi R_e s \ , \tag{5.30}$$

where R_e and R_i are the outside and inside diameter of the disc and s is the disc thickness; therefore:

$$V_{eff} = \pi(R_e^2 - R_i^2)s \ . \tag{5.31}$$

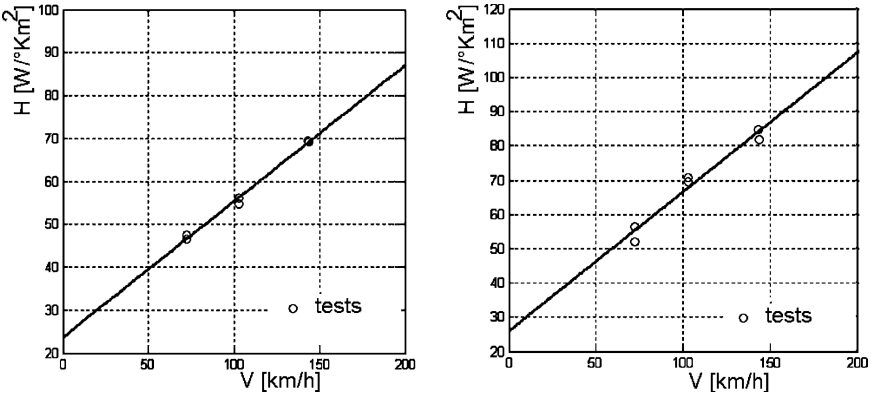


FIGURE 5.25. Diagram of the total convection coefficient H for rear discs (left) and front discs (right), as a function of vehicle speed.

Integrating the equation above, we can determine the diagram of the disc temperature as a function of time and of the convection coefficient.

This last could be identified by minimizing the difference between the results of experimental tests and calculations.

We can derive convection coefficient as a function of vehicle speed.

We have chosen for this example a linear correlation between convection coefficient and car speed.

Results are shown in Fig. 5.25.

It can be noticed that the front brake convection coefficient is greater than the rear, because the discs are ventilated and there is less covering effect from the body.

The assumptions about convection coefficient made for rotors cannot be applied to pads because the bulk of their absorbed heat cannot be dissipated to the air but is transferred to the calipers.

As a first approximation their convection coefficient has been set to the same value as the discs.

To perform a mono-dimensional analysis of the disc by the finite differences method, the disc has been divided into five layers; those closer to the pads are $\Delta x/2$ thick, while the three inner layers are Δx thick. The temperature will be uniquely a function of the distance x from the contact surface and of time.

The outer layers have a half thickness to factor in the higher temperature gradient near the contact with the pad.

Applying the first law of thermodynamics to each layer, we obtain a system of differential equations, whose solution allows us to identify temperatures for all layers at each time interval.

The thermal balance of the layers in contact with the pads is:

$$\begin{aligned}\rho c\tfrac{\Delta x}{2}S\tfrac{\partial T_i}{\partial t} = -H_a(S+\pi R_e\Delta x)(T_i-T_\infty) - \tfrac{4}{3}kS\tfrac{T_i-T_{i+1}}{\Delta x} + \\ -k_c\pi R_i(\Delta x - s_g\tfrac{\Delta x}{s})\tfrac{T_i-T_c}{\Delta x_c} + \tfrac{1}{2}\tfrac{\gamma}{\gamma+1}Q\ .\end{aligned} \tag{5.32}$$

In this equation:

γ	is the ratio of heat flux;
s_g	is the disc thickness, at the disc hub;
s	is the disc thickness, at the braking surface;
S	is the disc surface;
H_a	is the air convection coefficient;
T_c	is the hub temperature;
T_i	is the temperature of the layer i;
x_c	is the depth of the heat flux to the hub;
k_c	is the thermal conductivity between disc and hub.

For elements of thickness Δx is, instead:

$$\begin{aligned}\rho c \Delta x S \tfrac{\partial T_i}{\partial t} = H_a 2\pi R_e \Delta x (T_i - T_\infty) - kS \tfrac{T_i - T_{i+1}}{\Delta x} + \\ -q_{i,i-1} - k_c 2(\Delta x - s_g \tfrac{\Delta x}{s})\pi R_i \tfrac{T_i - T_c}{\Delta x} \ ,\end{aligned} \tag{5.33}$$

where $q_{i,i-1}$ is the heat flux that the layer i receives by conduction from the near layer; it can be written in a different way, depending on whether the element is closer to the outer or inner elements.

In the first case:

$$q_{i,i-1} = \frac{4}{3} kS \frac{T_i - T_{i-1}}{\Delta x} \ ; \tag{5.34}$$

in the second case:

$$q_{i,i-1} = kS \frac{T_i - T_{i-1}}{\Delta x} \ . \tag{5.35}$$

Because temperature distribution in a solid disc is symmetrical, the study can be limited, in this case, to half a disc, considering the symmetry plane as an adiabatic surface.

The calculation method has been applied to a car braking to a stop from 100 km/h, which takes about 3 s.

The result of the calculation is shown in Fig. 5.26. The disc is 10 mm thick and $s = 5$ mm is the middle of the disc; note that at the end of braking the temperature is almost uniform.

The temperature diagram as a function of time can be read at $s = 0$ mm and $s = 10$ mm, where it is identical for the assumptions that have been made.

A similar calculation can be made for the pad.

The hub and the caliper can be considered as concentrated masses; their temperature will therefore be a function of the time only.

Fading

During ideal repeated braking, the vehicle is slowed down from a higher to a lower speed; after it has slowed down it accelerates again to the initial speed and is then braked a number of times.

The temperature attained can be easily calculated in those conditions if braking power, cool down intervals and braking time are kept constant.

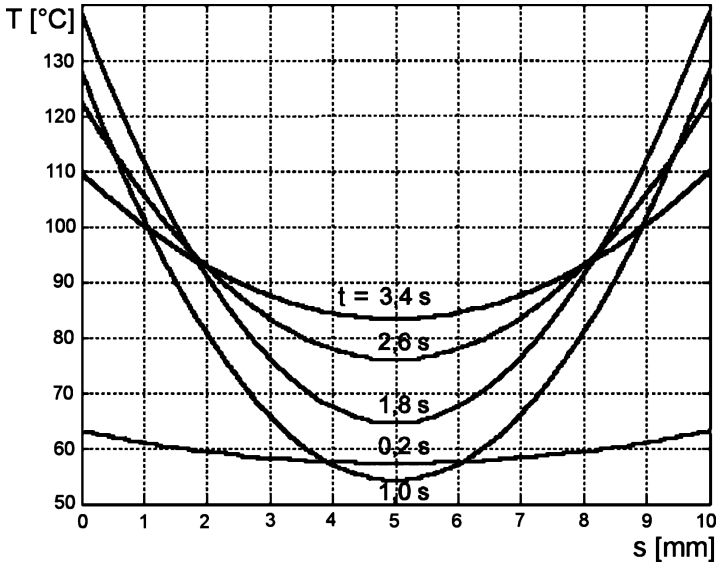


FIGURE 5.26. Diagrams of temperature T inside brake disc, as a function of time t and thickness s, in braking stop from 100 km/h.

In addition, rotors are modelled with concentrated parameter elements, with uniform temperature and heat transfer coefficients and properties that are kept constant.

If braking time were much shorter than cool down time, this last could be neglected.

In this case rotor temperature will increase uniformly according to the equation:

$$\Delta T = \frac{Qt}{\rho c V} \ , \tag{5.36}$$

where Q is the braking power, t is the braking time, c is the specific heat, V is the rotor volume and ρ is the material density.

To describe the behavior after a braking, the following differential equation is used:

$$\rho c V \frac{dT}{dt} = hA(T - T_\infty) \ , \tag{5.37}$$

where A is the rotor surface, h is the convection coefficient, T is the temperature at the time t and T_∞ is the ambient temperature.

Starting from an initial temperature T_0, the integration of the above equation leads to this solution:

$$\frac{T(t) - T_0}{T_0 - T_\infty} = e^{\frac{-hAt}{\rho c V}} \ . \tag{5.38}$$

Combining equations it is possible to calculate the temperatures at the second and following braking.

5.4.4 Test methods

As with other chassis components, brakes are also tested separately from the vehicle, to concentrate on particular aspects of their performance and to reduce development costs.

The test bench generally used is a dynamometer bench simulating vehicle inertia as well, as shown in Fig. 5.27.

The bench includes an electric motor 1, having the task of launching the flywheels 2 at different speeds, simulating the initial vehicle speed at which the braking begins.

The inertial mass of flywheels can be attached to the bench, to simulate the apparent rotating mass[2] of the vehicle pertinent to the wheel, where the brake is applied.

After the flywheels the brake rotor 3 is installed, while the fixed part of the brake (caliper or shoes) is bolted to the bench block.

A reaction torque meter measures the braking torque.

The simulation of braking includes working phases and rests: It is specified that temperatures must reach a defined value and, if braking is repeated, a predefined rest time must be imposed to simulate cool down; a ventilator may be necessary to obtain realistic results.

When the appropriate temperature is reached the rotation speed is adjusted to simulate the speed of the car and braking is performed.

Braking performance may be controlled at constant braking torque or at constant circuit pressure.

During braking many parameters are recorded such as speed, temperature, pressure, torque and the temperature of particular elements on which attention is focused.

Many test procedures are applied:

- Homologation regulation procedures

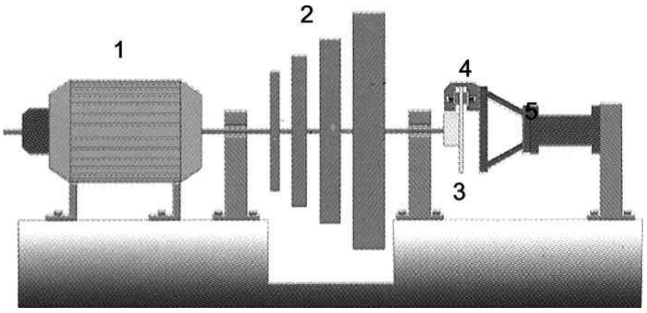


FIGURE 5.27. Scheme for testing a brake separated from its vehicle. The test bench includes a dynamometer and a set of flywheels simulating vehicle inertia.

[2]See the definition of apparent rotating mass in the second volume.

- EUROSPEC procedures, standardized between car and brake components manufacturers
- Procedures derived from specific company information or expertise

A particular object of bench test evaluation is mechanical and thermal rotor deformation; these measurements can be made without interrupting the test procedure, by using capacitive transducers that work without contact.

After a cycle of different tests simulating brake life on the car, linings and rotors wear can be measured to check their performance on the vehicle.

6
CONTROL SYSTEMS

This chapter is dedicated to control systems working on the steering mechanism, the brakes and the elastic and damping elements of suspensions; the sections are divided according to this classification.

Transmission control systems, working on the gearbox and differentials, will be examined in the second part.

This organization of subjects lays itself open to criticism, because the object of all these systems is often similar, even if they are applied to different parts of the chassis.

Most of these systems, in fact, are addressed to improving vehicle dynamic behavior through a suitable control of the forces exchanged between tires and ground; this control action can be the result of the breakdown of braking forces (through the brakes of each wheel), of the breakdown of vertical forces (through the elastic or damping elements of the suspension), of different steering angles and of the breakdown of driving forces (through the axle differential or the central differential in four wheel drive vehicles).

The following sections are primarily addressed to explaining how the cited chassis elements are modified to become actuators in a controlled system; these chapters will also outline the related control strategies, the rules that the control system must follow to obtain a result of improving the vehicle's dynamic behavior.

The study of the interaction of control systems with vehicle systems requires modeling the entire vehicle, in order to foresee its dynamic behavior and the modifications of vehicle functions. This study will be developed later, in the second volume.

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

6.1 STEERING CONTROL

A control system faster and more precise than an average driver, working directly on steering angles, can effectively stabilize and improve the dynamic behavior of the vehicle.

Steering angles calculated by the control system can be added to or subtracted from those imposed by the driver; in the first case the function obtained will be an improvement of maneuverability; in the second case, an improvement of stability or active safety results. The control system can modify the characteristics of the existing steering mechanism or can operate through a parallel mechanism on the rear wheels.

6.1.1 Rear wheel steering

RWS (*Rear Wheel Steering*) or 4WS (*Four Wheel Steering*) systems achieve additional steering of the rear wheels, as a function of different parameters; these can be front wheel steering angle, car speed, yaw velocity or lateral acceleration.

The schemes in Fig. 6.1 illustrate the purpose of steering the rear wheels, by using the kinematic steering plot we have introduced in the chapter on steering systems for front wheel steering (scheme a).

In all cases we consider kinematic steering only, at speeds low enough that sideslip angles can be neglected; plots outline the curvature radius R of the vehicle path.

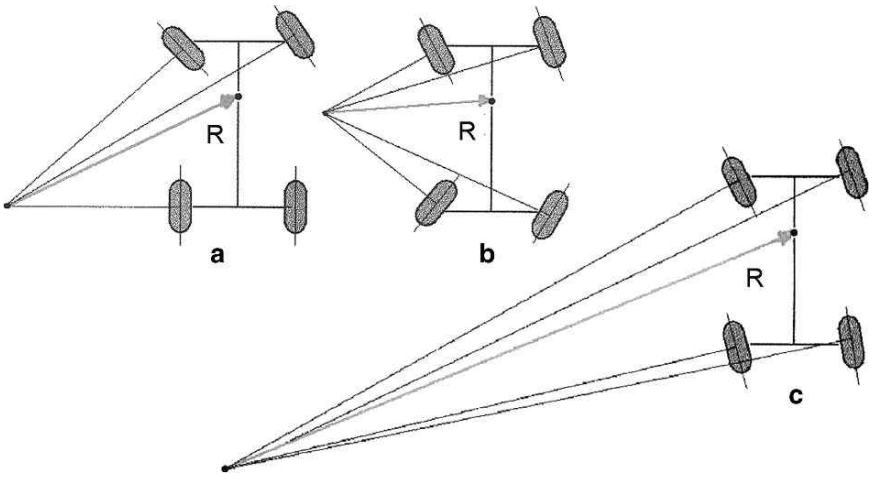


FIGURE 6.1. Kinematic steering of a vehicle with only front axle steering (scheme a) or with both axles steering in opposite directions (scheme b) or in the same direction (scheme c).

The addition of rear wheel steering angles allows better handling at low speed, thanks to the lower radius that can be achieved applying an opposite steering angle compared to the front wheels. With reference to scheme b it is possible to move the center of gravity of the car closer to the crossing point of the perpendiculars to the equator plane of the wheels.

In addition, it is possible to increase stability at high speeds, by applying a concordant steering angle to the rear wheels similar to that of the front wheels; yaw velocity can be decreased, with a decrease in vehicle sideslip angle. Theoretically it is possible to alter vehicle path with a pure translation motion with no yaw velocity, when the perpendiculars to the wheels are parallel, because all wheels have the same steering angle.

There are different categories of RWS systems, each of them having particular effects on vehicle dynamics:

- *Angle dependent*
- *Speed dependent*
- *Dynamic speed dependent*
- *Model following*

This classification is evidence of the evolution of electronic control systems in cars, more than a response to truly different needs.

Angle dependent

In an angle dependent system, rear steering angle is dependent solely on front wheel steering angle.

This condition can be obtained with a simple steering box, one that is mechanically connected to the front wheel steering box.

For small steering angles, as occurs at high vehicle speed, the rear wheels steer in the same direction, reducing vehicle sideslip and improving vehicle stability.

When the steering angle exceeds a certain threshold, typical of low speed maneuvers, rear wheel steering angle is opposite, to improve vehicle handling.

Typically the rear steering angle reversal occurs at about a 200 deg÷250 deg rotation of the steering wheel, when the rear steering angle is less than 2 deg in the same direction or 5 deg in the opposite. This open loop control strategy can be implemented with suitable mechanisms connected to the conventional steering mechanism.

An example of a purely mechanical four wheels steering mechanism appeared on the Honda Prelude in 1987. Figure 6.2 shows the effect of this steering mechanism on front and rear steering angles δ_1 and δ_2 as function of steering wheel angle δ_v.

Another example of an angle dependent rear wheel steering mechanism was patented in the same period by Fiat Research Center and is represented in Fig. 6.3.

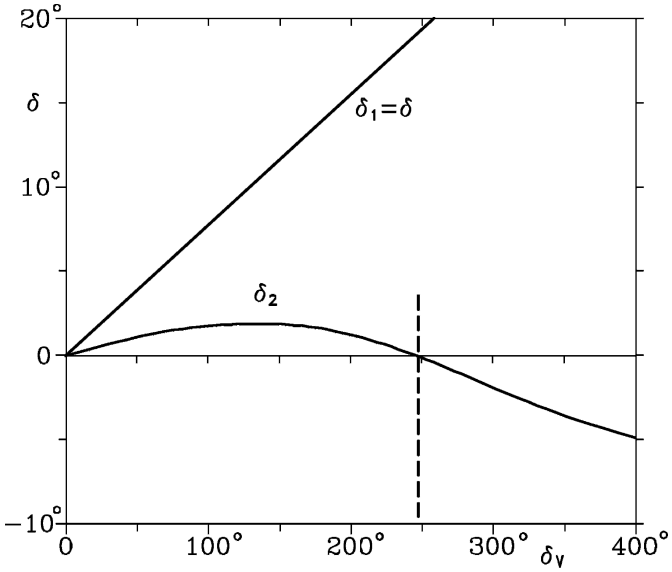


FIGURE 6.2. Steer angle on the front axle δ_1 and on the rear axle δ_2 as a function of the steering wheel angle δ_v.

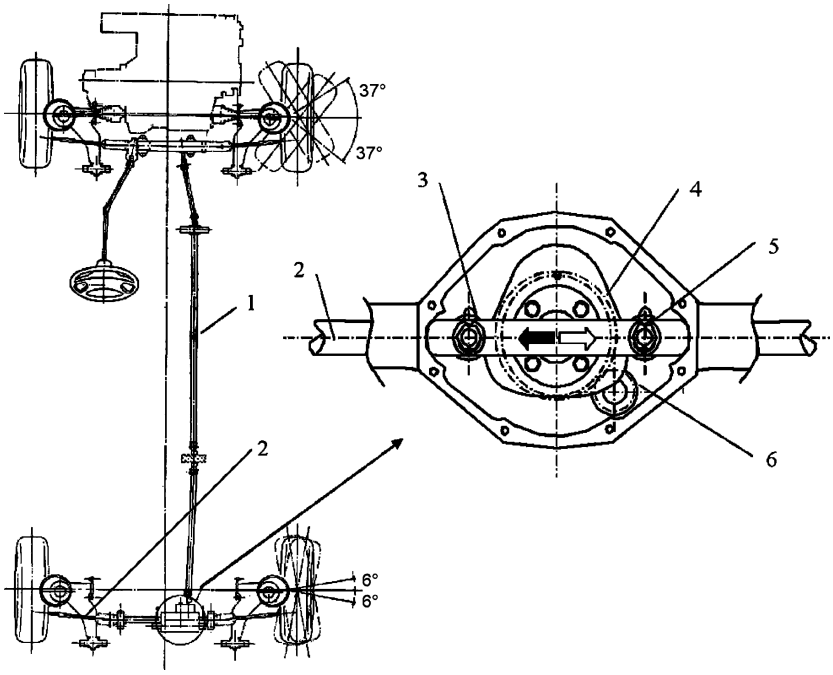


FIGURE 6.3. Mechanical control of rear wheel steering angle, angle dependant type. We can identify: Control shaft 1, rear steering push rods 2, roller contacts 3 and 5, steering cam 4 and steering gear 6 (CRF Patent).

Steering wheel rotation controls the rear axle steering box through a connection shaft and a gear 6. A suitable cam shifts two push rods 3 and 5 with a contact roller and, as a consequence, the two rear steering rods 2, to which they are connected. These steering rods act on the rear wheels.

The cam profile is such as to steer the rear axle in the same direction as the front axle, for small angles; for a clockwise steering angle (seen from above) on the front wheels, cam 4 rotates clockwise (in the enlargement at the upper right) shifting push rods to the left, with the lower profile, and then to the right with the higher profile.

There is a steering angle at which the rear wheels do not steer; the radial cam dimension is constant and equal to the distance between the rollers on the steering rods.

Speed dependent

A slightly more complex control strategy is the so-called speed dependent strategy; here the average rear wheel steering angle δ_2 can be defined with the formula:

$$\delta_2 = K(V)\ \delta_1\ , \tag{6.1}$$

where K is solely a function of the car speed V and δ_1 is the average front wheel steering angle.

As is obvious, a purely mechanical connection between steering axles is not viable without unacceptable complications (we can imagine, for instance, how car speed might be mechanically measured).

The most common objective of a speed dependent rear steering system is to make vehicle sideslip angle as low as possible. In such a case the control law can be derived by the mathematical models we will introduce in the second volume.

This is again an open loop approach making reference to average vehicle driving conditions.

Many applications have been introduced employing mechanical or hydraulic actuators that were subsequently discontinued.

A system of this kind was introduced on the Honda Prelude in 1992. In this case, rear wheel steering is achieved through an electric motor; sensors are installed on the car to measure steering angle and car speed. The control system maintains a quasi-static vehicle sideslip angle at all speeds.

A similar system is produced by Delphi (*Quadrasteer*) and is the only one still on the market.

Dynamic speed dependent

A more complete expression of a control law suitable for obtaining $\beta = 0$ in all conditions must have the following form:

$$\delta_2 = K(V)G(t)\ \delta_1\ , \tag{6.2}$$

where $G(t)$ is a suitable function of the time elapsed from the starting point of a steering input.

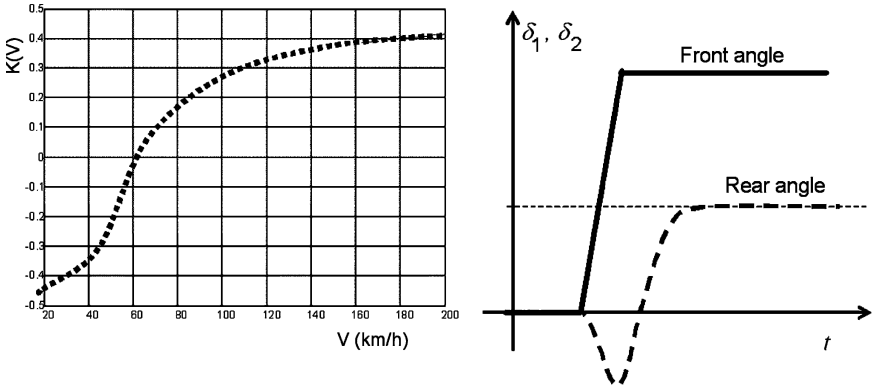


FIGURE 6.4. Diagram of the front and rear steering angles versus time in a steering transient with a dynamic speed dependent system.

A qualitative diagram of $K(V)$ is shown in Fig. 6.4. We see that at low speed (below 40 km/h) there is a negative gain, i.e. the rear steering angle is opposite to the front angle, to improve handling; at higher speeds the front wheels steer in the same direction to improve stability.

On the right side of the same figure a diagram of steering angle versus time is also shown. The steady state angle δ_r is modulated by the dynamic function $G(t)$ so as to obtain the response shown. In the first part of the diagram the opposite angle makes the onset of turning quicker while the following values are addressed to regime stabilization.

To reach this target the rear wheels are steered by a hydraulic actuator whose pressure is electronically controlled: The neutral position of the steering mechanism is defined by high stiffness return springs.

Model following

The control law calculating rear wheel steering angle includes two contributions.

An open loop control acts as the most sophisticated of the previously introduced strategies, while a closed loop control is finalized to compensate for errors detected between the actual and theoretical parameters predicted by a real time mathematical model; these parameters can be yaw speed and vehicle sideslip angle.

In this way the control system can react to situations that cannot be foreseen by a mathematical model such as:

- Changes in tires characteristics (wear, inflation pressure)
- Changes in vehicle mass and mass distribution
- Effect of external disturbances such as wind gusts (i.e. like when exiting a tunnel) or changes of road friction coefficient

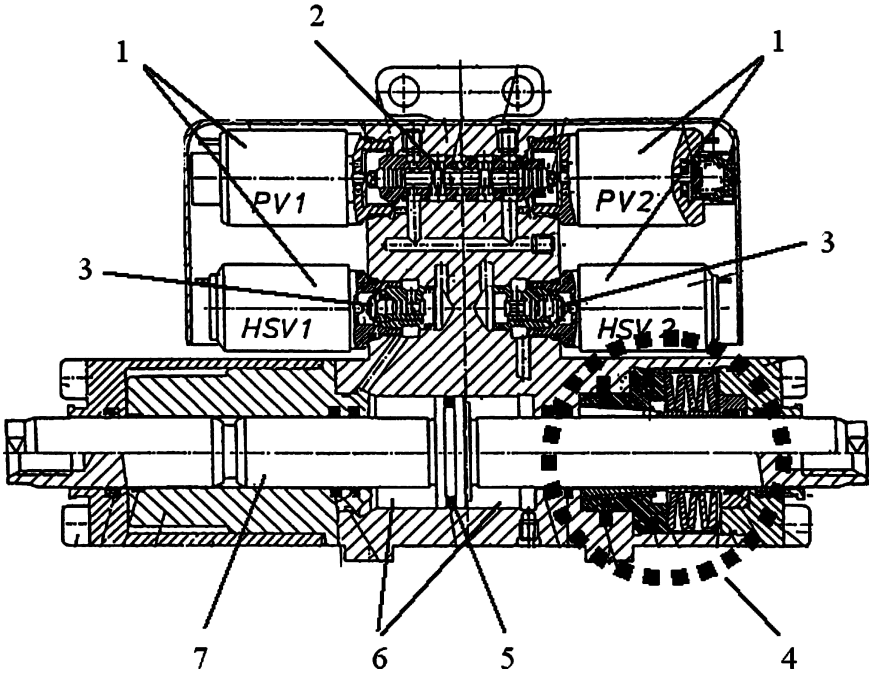


FIGURE 6.5. Rear steering angle actuator for a dynamic speed dependent control. Proportional valves 2 and on-off valves 3 are shown in the drawing, together with the safety brake 4, to keep the system blocked in case of failure of the pressure circuit (CRF Patent).

Response is always calibrated to the driving situation. Handling and stability are consequently improved.

Figure 6.5 shows a cross section of the rear wheel steering actuator, applied to a prototype developed by Fiat Research Center for a model using the RWS system.

This actuator includes two proportional electrovalves PV1 and PV2 to control the oil flow to the piston 5 to steer the rear wheels, and two on-off electrovalves (HSV1 and HSV2) to cut the oil flow to the cylinder chambers; a further electrovalve, not shown in the figure, opens the mechanical safety brake 4.

To move the actuator, all on-off valves must be open and the safety brake 4 released. The latter is made of a tapered bush around the steering rod 5. A set of bevel springs pushes the bush and keeps the brake naturally engaged; only oil pressure, if available, can disengage the brake.

It is important to outline the intrinsic safety feature of this system, required by regulations; the actuator guarantees steering angles to be fixed at the last actuation position; this feature alerts the driver, in the event of system failure.

Rear steering systems, after a period of relative interest during the 1990, at least on sports cars are now discontinued because of their higher cost, as

compared with VDC systems. Their marginal advantages in handling may induce manufacturers to reconsider this system for vehicles with restrictions on minimum turning radius, such as front wheel driven cars with large engines.

6.1.2 Variable ratio steering box

The variable ratio steering box represents an evolving technology in evolution, particularly for upper market high performance cars.

BMW, in cooperation with ZF and Bosch, has developed such a system with a continuously variable transmission ratio for the steering box.

Figure 6.6 represents, at right, the modified steering box; between steering column and pinion, a device made with an epicyclic gear 2 is applied, operated by an electric motor 1.

The scheme of the epicyclic gear is shown on the left of the same figure. The electric motor 1, controlled by an electronic circuit, acts through a worm gear V on the carrier P of an epicyclic gear 2, able to modify the transmission ratio between steering column i, connected to the steering wheel, and pinion u, geared with the rack 3.

As required by law, the system performs like a conventional steering box. In case of sudden failure, steering angle is not affected. In fact, if the electric motor stops, the worm gear gearbox, irreversible by design, keeps the carrier stopped in the last actuated position. The gear acts as a double reduction with planets at fixed position, while their rotation axis s is blocked, with a transmission ratio of 1:1 between sun gears S, without motion reversal.

An electronic control box is able to correct vehicle path as a function of yaw speed and lateral acceleration detected by sensors applied for the purpose.

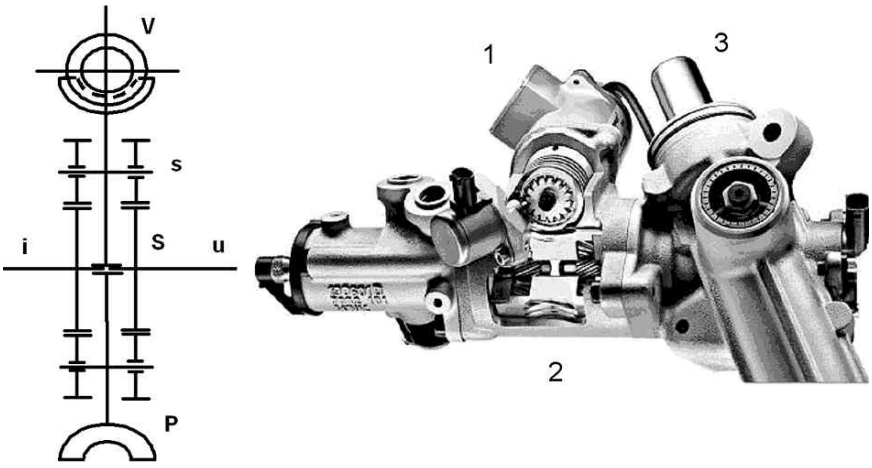


FIGURE 6.6. Actuator for front wheel steering control, made with an epicyclic gear adding to the steering wheel angle the contribution of an electric motor. At right the steering box; at left the kinematic scheme (BMW).

At low speed the electric motor contributes to reducing the effort required for steering to improve handling.

A different variable ratio steering box has been developed by Honda.

The regulation of the transmission ratio is performed by an electric actuator between the steering column and the pinion of the steering box.

At low ratio the actuator moves the pinion closer to the rack, to reduce its pitch radius; for a more direct ratio the rack is pulled away to increase pitch radius.

The control is able to change the steering wheel rotation angle from stop to stop from a minimum of 1.4 turn to a maximum of 2.4 turns.

6.1.3 Steer by wire

A *steer by wire* system, in which the mechanical connection between steering wheel and steering box is replaced by an electric or hydraulic transmission, is able to achieve all functions described in the previous paragraphs (auxiliary power, ratio variation, stability control) with a simpler mechanical system, at least from a conceptual standpoint.

The steering wheel drives a position sensor and is driven by a torque actuator able to replicate driving feed-back on the controls. The wheels are steered by an angle actuator under the command of the control system.

The steering wheel can be replaced by different controls, as, for example, joysticks or control sticks. The actuation power can be hydraulic or electric, with a preference for the last because of the easiness of interfacing electronic control circuits.

The expected advantages also include a better control layout with advantages for passive safety (reduced danger of contact with the driver after a crash), in roominess and driver seat adaptability; controls could be easily moved from one side of the vehicle to the other.

These systems could also be easily adapted to the specific needs of handicapped drivers.

This interesting conceptual simplification involves complications to guarantee the expected reliability; these systems are still not admitted by existing regulations that request a default positive mechanical drive between control and steering box in case of failure.

A system of this kind could easily integrate rear steering functions.

With two different and independent control inputs (δ_1, δ_2), an independent control of two degrees of freedom such as yaw speed and vehicle side slip angle is possible.

Additional potential advantages should be investigated by controlling each steering wheel separately on each axle.

The idea behind such controls should be to steer each wheel under kinematic conditions at all times, improving vehicle handling especially in narrow turns.

6.2 BRAKE CONTROL

Longitudinal forces generated by the braking system can affect not only longitudinal, but also lateral dynamics.

We suggest looking at Fig. 2.32, where the longitudinal friction coefficient μ_x is shown as a function of the longitudinal slip σ. In addition, the application of a longitudinal braking force F_x changes, at a given side slip angle α and vertical force F_z, lateral force F_y and self-aligning torque M_z, as shown in Figs. 2.60 and 2.61.

An empirical formula, commonly adopted to describe this phenomenon, has been proposed by Pacejika and commented upon in the chapter dedicated to the wheel.

Non-symmetric braking torques applied to the wheels can produce yaw moments capable of affecting vehicle path.

Brakes are therefore actuators not only suitable for changing vehicle speed, but also for controlling its path and traction capacity. The many control systems that are based upon this actuator are described in the following paragraphs, with Bosch trademarks and acronyms used for identification; these have entered into the common technical language.

6.2.1 ABS system

The objectives of ABS (*Anti-lock Braking System*) systems are, as already explained, to avoid excessive longitudinal slips that can reduce braking capacity and lateral dynamics control.

This objective must be reached for every possible value of friction coefficient (in principle, different for each wheel) of road slope and vehicle load.

In the mean time, the control system must minimize any disturbance resulting from sudden variation of the friction coefficient (such as driving on puddles, pits, manhole covers, etc.), suspension vibrations or tolerances in the fabrication or assembly process.

In addition, braking capacity should not be affected by transmission inertia (vehicle inertia at the driving wheels is affected by gear ratios and by the engagement or disengagement of the clutch). Finally, yaw torque disturbances caused by different friction coefficients on the wheels (the so called *μ-split conditions*), are to be reduced, as are vibrations on the brake pedal caused by braking torque regulation.

Input parameters of this control system are wheel speeds, measured by sensors on the wheel hubs, while the controlled parameter is the time derivative of the braking pressure at each brake. In fact, ABS controls are unable to set a given value for the braking pressure, but control its increase, decrease, or maintenance of existing pressure.

While simpler systems have been built in the past, we will refer solely to the four channels system, now universally applied; this system is characterized by a speed sensor and an independent pressure control for each wheel.

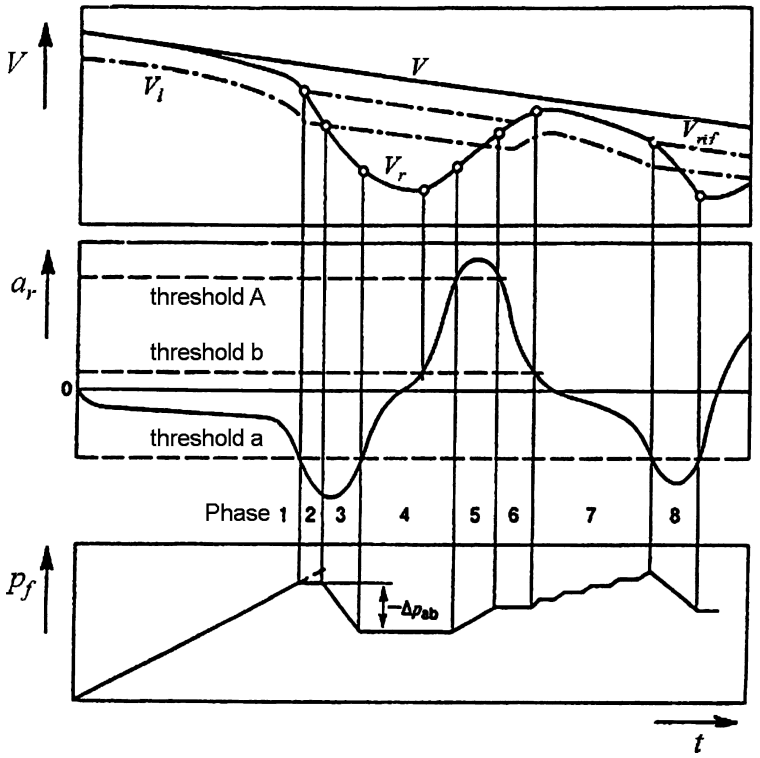


FIGURE 6.7. Diagrams of speed, acceleration and braking pressure in a braked wheel controlled by an ABS system.

To achieve this function, an ABS system must include:

- A master pump with power system, as used in traditional systems.
- A brake actuator for each wheel (disc or drum) as in traditional systems; we should notice that disc brakes are preferred to drum brakes, because they are simpler to regulate, being free of potential self-locking.
- A speed sensor for each wheel.
- A hydraulic pressure modulator, including an electronic control unit, regulation valves and a recirculation pump.

This components will be explained in a following paragraph; we will start now to explain the control strategy applied to fulfill the objectives.

Figure 6.7 presents a typical diagram of fundamental parameters for a wheel, during a progressive braking on a friction surface, with the intervention of the ABS system.

The first graph above shows the diagram versus time of:

- Vehicle speed V: This value is not measured by the control system, but is shown on the graph for reference.
- Reference vehicle speed V_{rif}: This parameter is estimated by the control system using algorithms that elaborate wheel speeds.
- Wheel peripheral speed V_r: This is the product of the wheel speed Ω_r, measured by sensors, and the rolling radius R_0.
- Limit speed V_l: This is the threshold of the wheel peripheral speed, below of which slip is too high to guarantee control targets.

The second diagram (in the middle) shows the peripheral acceleration of the wheel a_r, while the third, at the bottom, depicts the effective braking pressure on the actuator p_f as controlled by the electronic unit.

The electronic unit calculates the acceleration a_r, deriving the speed V_r measured by the wheel sensors; particular techniques are applied to filter signals from magnetic pick ups, thus suppressing noise.

Some characteristic threshold values are identified in the acceleration diagram; they are:

- Threshold a (value <0): This value is certainly higher than the maximum vehicle acceptable deceleration. It shows a marked difference between the actual braking moment and the maximum stable braking moment, due to excessive braking pressure or a sudden transient to the unstable area.
- Threshold A (value >0): This value shows a braking torque remarkably lower than the maximum stable braking torque; the corresponding slip has a small value, well below the maximum.
- Threshold b (value >0): Near to zero, this value underlines the return close to the maximum of the (F_x, σ) curve.

Seven characteristic phases of a braking can be identified, under the control of an ABS system.

Phase 1

The time $t = 0$ in the diagrams corresponds to the start of braking. Throughout phase 1 there is no intervention of the control system.

In fact, in this phase the wheel speed V_r is greater than the threshold V_l and the peripheral acceleration a_r is negative, in the same order of value as vehicle deceleration. The value of the latter is lower than the threshold a, that indicates the transition to the unstable field.

As a consequence, in a progressive braking, the piston pressure increases continuously, depending on driver actions and circuit delay. The acceleration will be maintained almost constant because the slip increase is such as to keep the friction torque increasing along with the braking torque.

Before reaching the threshold a, the acceleration will show a sudden change, underlining that the boundary with the unstable area is near.

Phase 2

As soon as threshold a is overtaken, phase 2 begins, where braking pressure is kept constant.

This occurs only after the threshold has been overtaken. In fact a sudden high speed braking, characterized by a high braking torque gradient, in combination with a tire characteristic curve not very steep in its stable part, implies a slow increase of the friction torque. Therefore it is possible to reach threshold a in stable condition and unnecessary to reduce the pressure in those conditions.

If the beginning of phase 2 represents the transition from stable to unstable area, the value of slip that has been reached is optimum, to be held constant during braking to produce good performance.

In this phase the wheel deceleration keeps decreasing.

This explains why the braking torque is constant while the slip increases, with reduction of friction torque (the limit of the F_x - σ curve has been overtaken).

Phase 3

Below the value V_l the slip is very high close to the wheel block; therefore crossing the corresponding threshold for the wheel speed V_r implies entrance into phase 3, characterized by decreasing pressure.

The consequent decrease of the braking torque makes an inversion of the acceleration diagram possible, which even below the threshold a begins to increase, while the wheel speed decreases, but with a reduced slope.

In phase 3 the wheel is still in the unstable area and the slip is still increasing. The friction torque decreases more slowly than the braking torque.

For this reason the acceleration begins to increase.

Phase 4

As soon as the acceleration overtakes the threshold a, we enter the phase 4, characterized by constant pressure. While in phase 1 crossing the threshold a means entrance into the unstable area, in this case crossing means return to the stable area, because in the meantime the braking torque has been stabilized at a lower value, corresponding to Δp_{ab}, in the third diagram.

This fact allows the acceleration to continue to increase until it is positive. As a consequence wheel speed V_r will decrease in the first part of this phase to a minimum and will then start to increase once again.

The growth of acceleration is due to the fact that, while braking torque is kept constant, friction torque grows until it is larger than the braking torque.

This growth occurs because of the slip decrease.

In the middle of this phase the deceleration again overtakes the threshold b. Should this not happen, it would be necessary to proceed to a further decrease of braking pressure.

This could happen because the pressure decrease in phase 3 has not been sufficient to allow the friction torque to become higher than the braking torque.

Phase 5

As soon as the acceleration overtakes the threshold A, phase 5 begins; it is characterized by a pressure increase.

In the first part of this phase, the longitudinal slip decrease causes an increase of the friction torque and a consequent increase of acceleration, until this increase changes the trend of the acceleration.

During this phase, the wheel speed continues to increase, because the acceleration is always positive.

Phase 6

Phase 6 begins as the acceleration overtakes, with negative slope, threshold A, where the pressure is kept constant. In this phase the wheel speed continues to increase and the slip decreases, but because the slip is again in the stable region, friction torque also decreases, together with acceleration, until this overtakes, always with negative slope, the threshold b.

Phase 7

At this time phase 7 begins, where the pressure in increased by steps to avoid a steep increment of the braking torque.

This kind of increase allows a better exploitation of the tires and decreases the influence of wheel inertia. At the beginning of this phase, wheel speed continues to increase and slip further decreases with the decrease of friction torque.

Acceleration continues to decrease to zero, returning then to negative values.

The value of zero corresponds to both torques being equal. As soon as the acceleration is negative, slip again begins to increase toward optimum values. These will be reached when the acceleration decreases below the threshold a, starting a new cycle, as we have described, beginning at phase 3.

We observe that, using ABS controls, the actuator pressure cannot ever exceed the pressure of the master pump. In other words the braking torque cannot override the action of the driver in control.

When braking on low friction grounds, some anti-lock systems recognize this situation and adopt slightly different strategies.

Other strategies are addressed to improving vehicle dynamic behavior.

For instance, the strategy *Select Low*, designed by Bosch, provides that the pressure of the rear actuators is determined by the wheel with lower friction, while the front wheels are controlled with different pressures.

In this strategy the yaw torque (consequent to different braking forces, for instance on μ-split grounds) is decreased, with a contained decrease in braking efficiency.

Another strategy, again by Bosch, brakes the front wheels using higher friction with a delayed pressure. In this case the yaw moment is only delayed, but the driver is given a better chance to compensate for this disturbance with a different steering angle. This delay should be calibrated to vehicle characteristics in order not to decrease braking performance excessively.

6.2.2 EBD system

The acronym EBD derives from *Electronic Brake Distributor*.

It is known that during braking, the inertia force applied to the center of gravity determines a vertical load transfer, increasing the vertical load of the front wheels and decreasing that of the rear wheels.

If a braking moment simply proportional to the vertical static load should be applied, the rear tires would loose their adhesion first, compromising vehicle path stability.

To avoid this outcome, a distributor valve has been introduced on braking circuits without ABS control. As we have described in a previous paragraph, this valve is addressed to limiting the rear wheel braking pressure.

This function can be performed by an ABS system too, and this added feature is called EBD.

For a given vehicle it is possible to calculate ideal pressure distributions, as shown in Fig. 5.12; this curve represents the locus of points on the p_{ant} - p_{post} plane that guarantees the maximum braking force on both axles.

This curve must be modified for each possible load combination on the vehicle. For instance the full load curve allows a higher rear braking than the empty condition, because load variations primarily affect rear wheel load.

EBD control tries to copy the ideal curve in a way superior to mechanical brake distributors.

To achieve this, vehicle mass is estimated by comparison of the master pump pressure with vehicle deceleration; the value of the mass can also be used to estimate mass breakdown.

Until control action on the brake pedal is minimal, the distributor follows the full load curve; additional slips of the rear axle are contained.

At higher master pump pressures, EBD control diversifies the pressure using ABS valves.

Note that such a strategy implies a precise calculation of slip, which depends upon vehicle speed. But this cannot be estimated precisely from wheel speeds. A control system maintaining the difference between rear and front wheels slip is simpler and similarly effective.

The components necessary to an EBD control system are the same as those on an ABS system; only software is affected, because specific algorithms are added for this function.

We should remember that an EBD system, as a mechanical distributor, is subjected to regulations on minimum deceleration in case of failure; the system must incorporate a recovery mode including a safe reduction of braking forces in the event of failure.

6.2.3 VDC system

VDC (*Vehicle Dynamics Control*) is also known through different commercial names, such as ESP (*Electronic Stability Program*) or VSC (*Vehicle Stability Control*).

The objective of such systems is to obtain a stable and predictable vehicle dynamic behavior always and to avoid tire operation over their stable limit. The actuator of this system is again wheel braking torque applied so as to obtain a suitable yaw moment on the vehicle.

Components necessary to a VDC control include:

- The master pump, with vacuum servo, as on a traditional system
- Brake actuators, as on a traditional system
- Wheel speed sensors, shared with ABS and EBD systems
- A hydraulic pressure modulator, including valves and load and recirculation pump, shared with ABS and EBD systems
- A steering wheel angle sensor
- Yaw speed and lateral acceleration sensors
- A connection line to the engine electronic control unit.

Steering wheel angle, yaw speed and lateral acceleration sensors are used to identify the primary parameters describing vehicle lateral dynamics.

The electric pump and valves are used for this system to generate pressure pulses for single brakes, independent of the driver's will.

When the vehicle is driven on a curve and the cornering force of one of the tires is less than that needed locally for equilibrium with the centrifugal force, the vehicle deviates from the previous path.

If this deficiency is localized in the front axle, the phenomenon is similar to understeering[1] and the vehicle describes a path with less curvature than in normal conditions; if it is localized in the rear axle, the phenomenon is similar to oversteering and the vehicle describes a path with higher curvature than in normal conditions.

The VDC control system tries to recover the initial path, with a correcting yaw torque compatible with tire limits.

---

[1]See definitions of understeering and oversteering in the second volume.

To do this, the electronic control unit calculates the expected yaw speed on the maneuver from steering angle and vehicle speed. If there are differences from the measured yaw speed, four braking torques are calculated that are useful for correcting this situation or at least reducing vehicle speed, if friction limits are surpassed.

A reduction in engine torque contributes also to this speed reduction; this is the purpose of the communication channel between engine control and VDC control systems.

Engine torque in reduced by spark advance reduction, injected fuel reduction or throttle valve reduction in engines driven by wire.

These interventions must not induce slips that bring the tire into the region of side slip or longitudinal slip instability.

Two control loops are necessary. One calculates the braking forces necessary to stabilization; an innermost loop controls wheel slip so as to obtain the desired longitudinal and lateral forces.

A mathematical model must be suitable for predicting vehicle motion consequent to load variation, at different vehicle load, road friction and tire wear.

6.2.4 ASR system

This acronym derives from *Anti-spin Regulator.*

Critical driving situations are not solely the domain of braking or curves. Accelerations and traction forces may also cause vehicle instability.

ASR systems correct for these situations by avoiding excessive slips on the driving wheels.

This result is achieved either by temporarily reducing the torque applied by the engine to the driving wheels or by applying a braking torque from the control system capable of generating torques autonomously.

ASR, as stated earlier, is an additional function of ABS systems, implying communication with the engine electronic control unit.

The ASR control unit is integrated with the ABS control and shares with it a number of components, such as wheel speed sensors and pressure control valves.

The components necessary to an ASR system are:

- Brake actuators, as for a traditional circuit
- Wheel sped sensors, shared with ABS and EBD systems
- A hydraulic pressure modulator, including valves and load and recirculation pump, shared with ABS and EBD systems
- A communication line with the engine control unit.

If throttle valve and brakes are available as actuators, the best performance can be obtained; because most engine controls are now driven by wire, this option no longer presents difficulties.

The control strategy can be outlined according to five phases.

Phase 1

The driving wheel longitudinal slip is calculated by comparing their speeds with idle wheel speeds.

Phase 2

The beginning of this phase is characterized by the longitudinal slip of one of the driving wheels surpassing a limit threshold; a suitable braking force is calculated to reduce this slip and consequently the driving force.

Phase 3

The slip is again below the safety threshold and the braking pressure is decreased.

Phase 4

Phase 4 is similar to phase 1, but begins if an excessive slip is also shown by the other wheel.

Phase 5

During this phase, if all driving wheels were braked, the engine torque would also be reduced so as to limit longitudinal slip.

ASR systems can be implemented by the MSR system (*Motor Spin Regulator*), which provides for reducing the engine braking torque during gear box down shifts or at sudden accelerator pedal release.

A potential slip situation could occur on low friction grounds (snow, mud, etc.) in low gears. In this case MSR systems require an increment of the engine torque, so as to reduce the engine braking effect.

6.2.5 BAS system

BAS systems (*Brake Assist System*) have the task of applying a constant full braking pressure, independent of the driver's will, in an attempt to decrease the stopping distance as much as possible, in emergency situations.

Such control systems are useful in panic braking by non-expert drivers; moreover expert drivers not suited to ABS systems may instinctively decrease the force on the brake pedal, trying to avoid slips that the control system will in any case avoid. For impaired drivers, the pedal force may be inadequate to exploit the full braking capacity.

Two alternative strategies may be applied:

- When a panic braking situation is detected, the system generates the maximum braking force independently of the driver's will.

- When a panic situation is detected, a supplementary braking force is applied, modulated by the driver's action on the brake control.

A panic situation is usually detected when the braking pressure time gradient exceeds a threshold value.

In the second strategy the deceleration is modulated by the force and speed of the driver's input. An ampler field of intervention is available than in the first strategy.

The components necessary to a BAS system are again the same as for the ABS system, once the hydraulic circuit has been modified with a suitable storage capacity.

The pressure modulator maintains a certain quantity of high pressure oil in a reservoir. The BAS system can apply a larger pressure as determined by the master pump.

6.2.6 Brake control hardware components

All brake control systems apply the following sensors and actuators.

Sensors

Wheel speed sensors are toothed wheels with magnetic pick-ups.

The toothed wheel RF in Fig. 6.8 has a rectangular profile made by ferromagnetic material, facing a magnetic pick-up; wheel rotation changes the magnetic flux captured by the pick-ups.

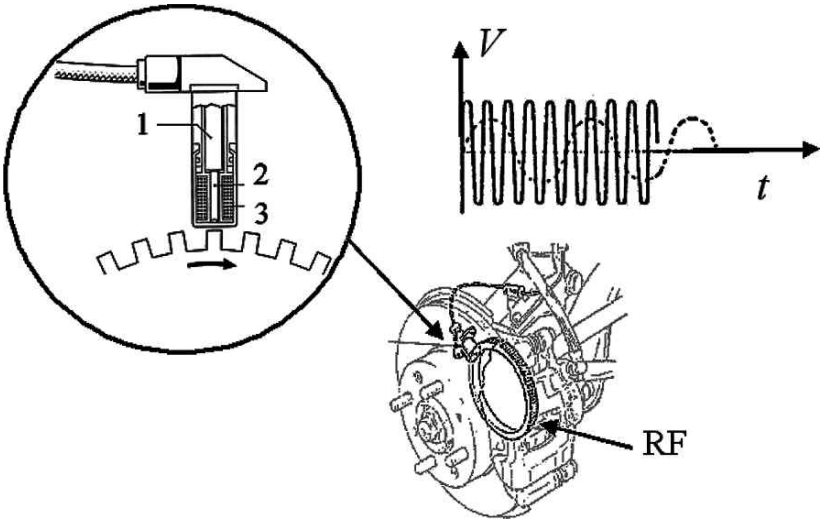


FIGURE 6.8. Installation scheme for a wheel speed sensor; RF is the toothed wheel. The enlargement shows the detail of the magnetic pick-up; at right are shown two diagrams of the generated voltage at low speed (dotted line) and high speed (solid line).

Pick-ups are made by a magnetic core 1 and a soft iron core 2. The magnetic field created by the magnetic core is periodically altered by the profile motion of the wheel and its value is measured by the coil 3, which will supply a voltage where period and amplitude are determined by rotational speed.

On the diagram at the top right the voltage is shown qualitatively as a function of time, for a low revolution speed (dotted line) and a high revolution speed (solid line). This signal is amplified and squared by an electronic circuit; the peak count supplies the requested speed value.

In current wheel speed sensors, electronic amplifier and signal conditioner are integrated into the magnetic pick-up, which can be interfaced directly with the control microprocessor.

The yaw speed sensor (gyrometer) and the acceleration sensor (accelerometer) are usually integrated into a single unit, including a solid state electronic circuit.

The working principle of the accelerometer is simple and consists in measuring deformations of a small structure subject to vehicle inertia forces. This measurement can be made by piezoelectricity, where an elastic deformation is converted to a voltage variation.

Yaw speed is also converted to an acceleration; it is measured by the Coriolis acceleration of the prongs of a diapason, brought to a forced oscillation by external means.

Steering wheel angle is measured of a potentiometer on the steering column or by an optical encoder. The requested precision is in the range of a tenth of a degree.

The encoder is still made by a toothed wheel, interrupting periodically the light beam received by a photo diode; the voltage pulse is measured and counted by an electronic circuit.

At the state of the art all these signals, converted to digital units, are made available to the CAN network of the vehicle, which makes them available to the brake control unit and to other services, as for instance, an electric power steering system.

Actuators

Figure 6.9 illustrates the functional scheme of a hydraulic modulator for an electronic brake control system.

Two independent hydraulic circuits can be identified, in this case, according to the X scheme, required to satisfy regulations on emergency circuit.

Six electrovalves normally open (ENO1..ENO6) are present in the circuit and six electrovalves normally closed (ENC1..ENC6). Normally open means that the hydraulic circuit is open when the coil is switched-off. The circuit also includes two pressure accumulators, along with two electric pumps used to speed up pressure decrease by ABS and VDC modulation and to create pressure during VDC and ASR modulation.

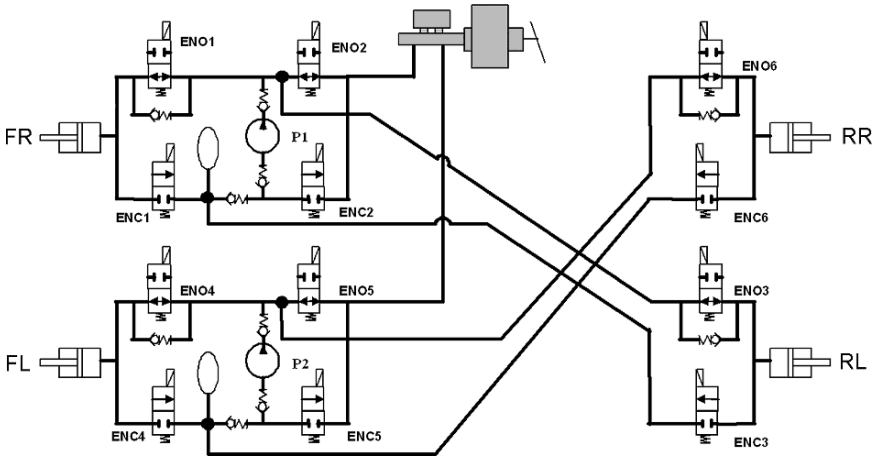


FIGURE 6.9. Functional scheme of a hydraulic brake modulator for a four wheel vehicle. ENO valves are normally open; ENC are normally closed.

These pumps are made of two radial pistons moved by a cam and an electric motor.

We make reference to the front right wheel (FR) only; the default configuration, when no intervention is required to the control system and all electrovalves are switched off, is represented in the figure. There is a direct connection between the master pump and the brake actuator FR.

This condition also allows conventional operation of the hydraulic circuit in case of control system failure.

When it is necessary to maintain a constant pressure in the brake actuator, independent of the driver's will, the electrovalve ENO1 is switched on and isolates the actuator from the rest of the circuit.

If it becomes necessary to decrease the pressure, the electrovalve ENC1 will also be switched on to allow that portion of the oil in the actuator cylinder to return into the master cylinder atmospheric pressure reservoir, with the help of the electric pump, if necessary.

To increase the pressure independent of the pressure in the master pump (VDC and ASR regulation), the electrovalve ENC2 is switched on putting the pump P1 in communication with the reservoir on the master pump; in the mean time the electrovalve ENO2 is switched on to avoid pressure oil return into the reservoir.

6.2.7 Hybrid and electrohydraulic circuits

The hydraulic braking systems we have examined until now reach a high level of performance but have also become quite complex and expensive.

The transmission of force from the pedal to the actuators is made using the hydraulic pressure; the need to contain this force within limits that are

ergonomically compatible with the average driver has justified the addition of a vacuum actuator.

The vacuum pressure, sometimes not sufficient or unavailable, as in diesel engines, has required the introduction of electric vacuum pumps.

Control force generation alone requires three different energy vectors.

Control systems require oil in pressure and electric pumps, involving two additional energy vectors. Electric energy looks to be the primary source for control force generation and regulation.

This irrational situation is justified by historic and economic reasons, because each of the above devices was born as an optional accessory, to be implemented into a conventional system, and the high volume of each of these components has discouraged the development of different, even more rational, solutions.

In some cases regulations have also required redundancies for reliability and safety, contributing to maintaining this situation.

The diffusion of power brakes and ABS, now at 100% on new cars, encourages the development of new system architectures, addressed to avoiding vacuum energy by using hydraulic power sources only or avoiding any other kind of energy, but the electric used for actuators and regulation.

Even if not largely present in series production, some vehicles feature so called hybrid systems that have a state of the art circuit on the front wheels and electric brakes on the rear wheels.

Hydraulic elements are eliminated for rear wheels only, on actuators and regulating circuits. The cost advantage is limited, probably attractive for electric parking brake applications only, but this could be the starting point of the development of an interesting new technology.

Figure 6.10 shows the rear brake actuator for a hybrid system.

This system is completed by an electronic control for electric motors and a pressure sensor to interpret the driver's intentions from master pump pressure; this sensor is already present in VDC control systems.

To produce an adequate clamping force on the rear pads with a small size electric motor, a reduction gear with high transmission ratio is adopted; in our case a double stage gearbox is used, including an epicyclic gear and a worm gear with ball recirculation. This feature is necessary for reducing friction and guaranteeing a quick release of the braking force.

A position sensor, not shown in this figure, is used to release the braking force, thus maintaining the clearance between pads and disc at a preset value, independent of wear.

An electromagnetic brake 8 is actuated with inverted logics; when the electric motor is off, the electric brake is on, stopping the motor and avoiding motion of the pads. This condition occurs in a stopped vehicle, held by the parking brake, or in the event of failures. The maintenance of the state of the braking system before the failure is judged to be the best policy. When the electric motor is on, the electric brake is off, leaving the pads free to move.

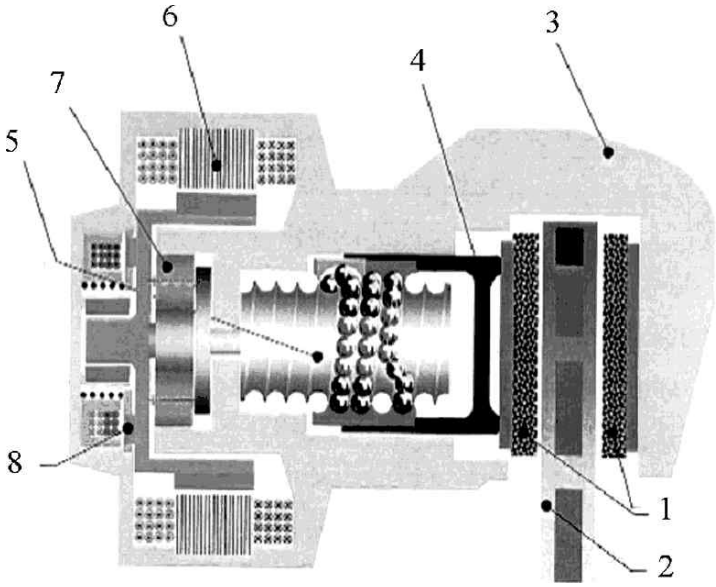


FIGURE 6.10. Cross section of an electric actuator for rear brakes, also used for the parking function. We can see: 1: disc, 2: linings, 3: caliper, 4: actuator cylinder, 5: ball recirculating screw, 6: electric motor, 7: epicyclic reduction gear, 8: brake to hold the actuator in position.

This kind of brake allows a further brake control system called EPB (*Electric Parking Brake*) to be implemented.

The purpose of this control includes power assisted control of the parking brake (no energy is requested of the driver) and the HH (*Hill Holding*) function, where parking brakes are actuated automatically when the vehicle is stopped on a slope. The stopping condition is identified by having the brake pedal depressed once, the accelerator pedal released and the car speed kept at zero. At the next start up, the parking brake is released only when the driving wheels are able to move the car, without undesired backward motion of the car.

Hybrid systems and electrohydraulic brakes, which we will introduce later, are the precursor of remote braking systems (*brake by wire*), that will probably be developed in coming years.

Electrohydraulic brakes (EHB, *Electronic Hydraulic Brakes*) also include in a conventional electric control system the function of power amplification.

Brake actuators remain the same, but the master pump is without vacuum actuator, no longer necessary; the brake pressure is generated by the electric pump only, while the master pump guarantees the residual braking capacity of the emergency brake.

The pedal feedback is given by regulating the pressure from a pressure accumulator.

The signal from the pedal to the electronic control unit flows through two parallel redundant channels: A pedal position sensor and a pressure sensor at the outlet of the master pump.

For pressure regulation only proportional valves are used. These regulate the pressure coming from an accumulator according to the pressure on the pedal.

Two on-off valves cut the circuit of the master pump; these valves can put pressure on the circuit when the pedal is not depressed for the VDC and ASR functions.

The ultimate goal will be a fully electric system with no hydraulic energy contribution, the so called EMB (*Electro Mechanical Brakes*).

The actuators could be similar to those we have presented in Fig. 6.10. There will be no positive connections between the pedal and the actuators.

This kind of system will require the development of redundancies capable of obtaining a level of safety and reliability at least equal to that of present systems; this objective will affect not only sensor and actuator design, but also the communication network.

6.3 SUSPENSION CONTROL

Suspension systems must satisfy conflicting objectives in terms of comfort and active safety. In fact, as will be explained in the second volume, suspensions able to guarantee a high level of comfort should be soft and poorly damped, while to guarantee a constant contact of wheels with the ground they should be rigid and damped.

Suspension systems made by springs and dampers are also called *passive*, because they react to forces coming from the road with energy contributions that are solely negative; they can, in fact, only waste energy.

Passive suspensions are designed to offer a reasonable compromise between comfort and passive safety, considering the character of the vehicle in question.

The characteristics of vehicle suspensions depend in fact on precise design options. Sports cars have stiff suspensions, not suitable for absorbing significant road unevenness, but finalized to a superior stability, under high cornering forces as well; normal cars instead feature softer suspensions, which allow higher comfort with some limitation on dynamic performance.

The limit of these passive suspensions can be easily explained by the impossibility of managing two independent parameters – body vertical accelerations (related to comfort) and vertical force variations (related to active safety) – with a single parameter, the suspension damping coefficient. The two objectives are independent and their optimum values are obtained with different damping coefficients.

At least conceptually, the best compromise can be obtained by adapting the suspension damping coefficient to the most important priority, which changes from time to time.

Tangible improvements have been obtained by adopting suspension components capable of adapting their mechanic characteristics to the changing needs of different driving situations.

Applying micro electronics and improved mechanical components, passive suspensions have evolved into *adaptive* or *controlled* suspensions; they need positive energy contributions from the outside. If these contribution are important they are called also *active* suspensions.

For controlled suspensions a classification is made considering the intensity of the energy contribution:

- *Semi-active* suspensions have an energy consumption limited by static vehicle trim control or damping control; the energy contribution is not sufficient to modify the vehicle trim in a time consistent with the oscillation period of the suspensions.

- *Active* suspensions have a significant energy contribution provided by dynamic vehicle trim control; the energy contribution is sufficient to modify vehicle trim in a time consistent with the oscillation period of the suspension; the control system is able to keep vehicle trim unchanged at almost any time, on any road.

Both types of suspension should include the following components:

- An actuator suitable for applying a force depending on what is calculated by the control system. Sometimes electric valves are used that supply a defined fluid flow, corresponding to a speed of the actuator; therefore the actuator speed, more than actuator force is controlled.

- Sensors measuring the significant parameters of vehicle kinematics.

- An electronic control system.

- A power plant, feeding actuators through valves; this plant is always present at different sizes, according to the system's active or passive features.

Considering the objective of the suspension control system, the following alternative may be considered:

- *Trim control*: A quasi-static control provides for a constant vertical static displacement of the rear axle or of both axles, at any vehicle load.

- *Damping control*: The damping coefficient of the shock absorber is adapted to different situations; static displacements are not affected.

- *Roll control*: Vehicle roll and roll speed are limited dynamically, as compared with conventional systems.

- *Full active control*: All above objectives are pursued in any dynamic situation.

The energy requested by the control system is significant in the third system and maximum in the fourth.

6.3.1 Trim control

Rear axle trim controls correct the height of the rear axle by decreasing the natural compression stroke when payloads increase. This effect can be obtained using hydraulic and pneumatic actuators.

Figure 6.11 shows the layout and functional scheme of this system on a car. This system includes an oil reservoir 1, at atmospheric pressure, a pump 2 driven by the engine, a pressure or trim regulator 3, connecting the two hydraulic actuators 5 with the pressure oil accumulator 4.

If the vehicle load has been changed, the rear suspension stroke variation is perceived when the vehicle begins its next trip, and some oil is sent to the actuators in order to obtain the desired trim. Hydraulic actuators work in series with suspension springs and compensate for spring compression.

A pneumatic system can be easily made by using elastic elements, similar to those introduced in Fig. 3.59; in this case, springs and actuators are integrated in a single component, the air spring. The control system can change the quantity of air in the elastic element, changing the pressure and obtaining the same position of equilibrium with an increased load.

It should be noticed that pneumatic systems change suspension flexibility too, as explained in the chapter dedicated to suspensions.

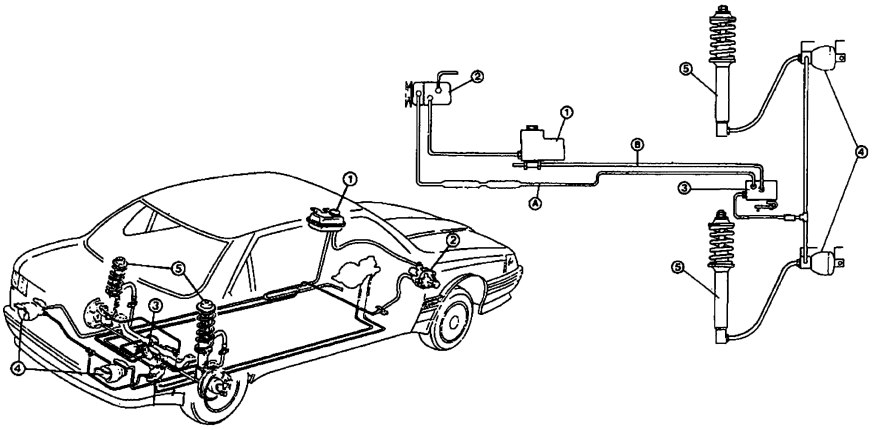


FIGURE 6.11. Hydraulic rear axle trim control; we can identify: 1: reservoir, 2: oil pump, 3: regulator, 4: accumulators, 5: hydraulic actuators (FIAT).

Another solution, quite popular on cars, is the Nivomat system, developed by ZF. It is made by single integrated units that replace rear shock absorbers; these integrate pump, actuator and regulator.

The energy necessary to raise the suspension after a load increase is derived from the suspension strokes on an uneven road. Figure 6.12 shows a section of this unit and its functional scheme.

In a single unit, two chambers are built: The low pressure chamber 1, at the bottom, and the high pressure chamber 2, at the top; this chamber is divided into two sections by the diaphragm 3. The outside section is filled with compressed gas, while the inside section is filled with oil. In an empty vehicle the two chambers are at the same pressure, about 25 bar; at full load, the low pressure chamber is at about 8 bar, while the high pressure chamber is at about 80 bar.

Inside the unit there is a conventional shock absorber 4, whose rod 5 is connected with the piston 6, bearing the choke valves. A drilled rod 7, at the bottom of the unit, moves inside a hollow cylinder 8 and works as an oil pump. On the outer surface of the rod 7 a hole 9 acts as a trim (vertical suspension displacement) sensor.

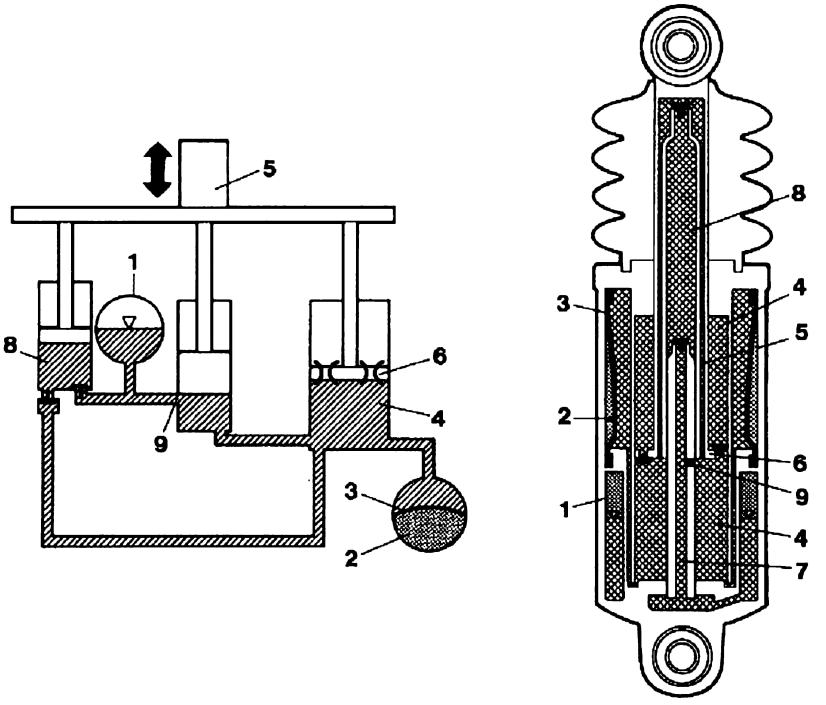


FIGURE 6.12. Cross section of a Nivomat unit (at right) and its functional scheme (at left). In this scheme, we can see the low pressure chamber 1, the high pressure chamber 2, and the shock absorber 4 (ZF).

Because of the vertical suspension stroke, produced by uneven ground, the oil in the low pressure chamber is aspirated by the pump, through the drilled rod 7. The pump sends the oil through the lower chamber of the shock absorber 4, into the high pressure chamber 4, compressing the gas separated by the diaphragm.

The higher pressure thus built up works on chamber 4, extending the rod 5 and raising the vehicle. When the hole 9 exits the cylinder, the desired trim is reached and the oil flows to the low pressure chamber, through the driller rod 7.

Some systems make front and rear axle trim control possible.

These are essentially two hydraulic or pneumatic systems working on each axle.

A characteristic system of this kind has been developed and adopted by Citroën. It uses a hydro pneumatic suspension, which combines air springs with high pressure gas (nitrogen), along with the original shock absorbers, integrated into a single hydraulic circuit.

6.3.2 Damping control

The idea behind damping control systems is to apply a damping coefficient to shock absorbers that can be adjusted to different values; the adjustment can be continuous or made at discrete levels. The damping coefficient is modified as a function of vehicle speed and vertical acceleration, with strategies that can be simple or sophisticated.

The scheme in Fig. 6.13 shows a shock absorber with a solenoid valve used to change the damping coefficient.

By controlling the current into the solenoid 2, it is possible to change the position of the anchor 3, through the reaction of the spring 4; the position of the anchor determines the passage orifice and, therefore, the damping coefficient of the shock absorber.

Controlled dampers can be classified according to the type of control strategies as:

- Adaptive
- Semi-active

The operating principle of an adaptive control typically includes the following sensors:

- Vertical accelerometers applied to the vehicle body, at least 1, with a maximum of 3
- A lateral accelerometer on the front of the vehicle or, alternatively, a steering angle sensor
- A pressure on-off switch on the braking circuit
- A car speed sensor (may be the same as used in the instrument cluster or ABS system)

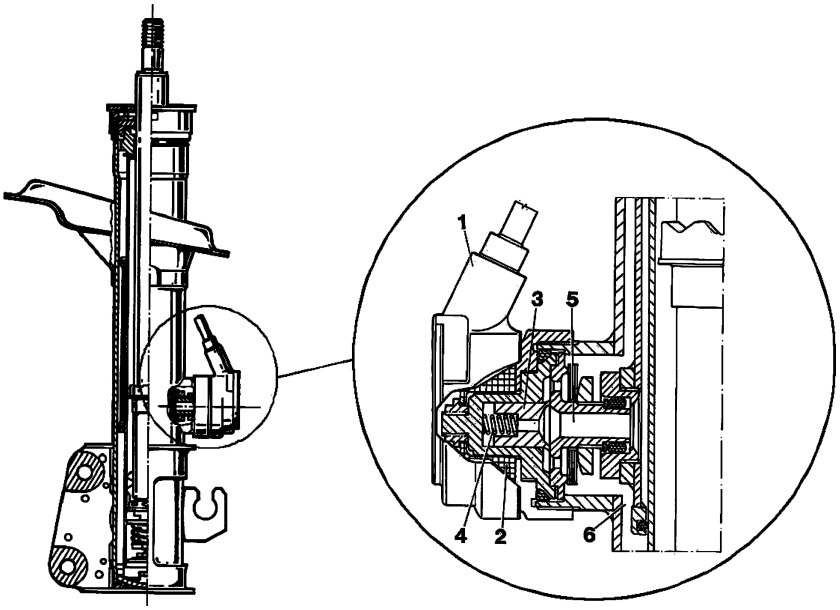


FIGURE 6.13. Adjustable shock absorber; a solenoid valve (magnified at right) can change the choke orifice dimension and, consequently, the damping coefficient.

At very low speeds the damping coefficient is set to high levels so as to limit car bounce during parking maneuvers.

To improve comfort, the damping coefficient is reduced very rapidly at higher speeds and kept at low values until about 120 km/h; at higher speeds it is again increased to improve vehicle stability (high speed roads are, usually, even).

The longitudinal acceleration sensor is used to identify medium speed sport driving, characterized by the frequent use of brakes and accelerator. Simple logic algorithms elaborating these two parameters are used to increase the damping coefficient, in order to reduce pitch angles. The detection of lateral acceleration over certain thresholds can further increase the damping coefficient to reduce roll angles and better stabilize the vehicle in turns.

Semi-active systems applied to vehicles are more complex but enable better control of body motion and improved performance; they are, usually, based upon the *sky-hook* damping theory.

According to this theory, the ideal situation is reached when the vehicle body is connected to the inertial reference system (the sky) by a shock absorber.

Because this architecture is not feasible, the control system is able to calculate a damping force proportional to the absolute vertical speed of the suspension, taking into account body and wheel speeds, with the formula:

$$F_{za,i} = C_s \dot{Z}_G + C(\dot{z}_i - \dot{Z}_G) \ , \tag{6.3}$$

where:

$F_{za,i}$ is the force exerted by the shock absorber of the wheel i,
$\dot{Z}_G$ is the absolute vertical speed of the vehicle center of gravity,
$\dot{z}_i$ is the relative vertical speed of the suspension of the wheel i,
C_s is the sky-hook damping coefficient,
C is the damping coefficient of the suspension of the wheel i.

The body speed is calculated by integrating the signals of three accelerometers on the car body, measuring the three components of the vehicle reference system. Relative suspension speeds are calculated by measuring the relative displacements of shock absorbers.

When the directions of absolute and relative speeds are opposite and their difference is significant, the ideal damping force $F_{za,i}$, to be applied, could have an opposite direction from the compression or extension speed of the actuator. In this case a full application of the sky-hook theory would mean applying double effect actuators with an energy contribution from the outside. This fact would complicate the system considerably.

To avoid this problem, a simplified theory has been developed using conventional adjustable shock absorbers (semi-active system).

An adjustable shock absorber allows all conditions to be satisfied, so that the requested force has the same direction as the relative suspension speed. In other situations, where force and relative speed are opposite, the force is set as low as possible, in order to minimize errors between the desired and the actual value.

The results obtained with these techniques are quite interesting and significantly superior to those obtained with adaptive techniques.

New components are appearing on the market, based upon a new technology; these are the magneto-rheologic shock absorbers.

Magneto-rheologic fluids modify their viscosity when a magnetic field is applied. They are made from a suspension of ferromagnetic particles in a mineral or synthetic oil; additives are provided to reduce wear in restricted orifices and maintain suspension.

Particles in suspension represent a fraction between 20% and 60% of the total volume. With no magnetic field, the fluid behaves in Newtonian fashion; we have:

$$\tau = \nu\gamma \,, \tag{6.4}$$

where:

τ is the shear stress,
ν is the kinematic viscosity,
γ is the deformation speed.

When a magnetic field is applied, particles tend to aggregate in chains along the field lines and the fluid can be modelled according to Bingham; the shear is given by the formula:

$$\tau = \tau_0 + \nu\gamma \,, \tag{6.5}$$

where τ_0 is the limit shear induced by the magnetic field.

Under the action of the magnetic field, the particle aggregation makes the fluid similar to a solid body, until the shear stress is below a limit value τ_0; if this value is exceeded, chains begin to break and the behavior is again modelled by Newton's formula.

This limit value is controlled by the magnetic field intensity. The magneto-rheologic shock absorbers have reduced response time, on the order of $5 - 8$ ms, that allow quicker control response, but demand faster elaboration times to take full advantage of this feature.

Magneto-rheologic components are conceptually simpler than their electro hydraulic counterparts; on the other hand, they must have a stiffer structure to avoid damage caused by the fluid when in solid state.

6.3.3 Roll control

A first classification of roll control systems should consider the type of actuator used.

These actuators can be set on the anti-roll bar, at its middle or at one end; the choice is determined by installation issues.

A second alternative is to intervene in the elastic element flexibility solely through roll motion; this is possible with hydro pneumatic suspensions.

Figure 6.14 shows an example of an actuator set in the middle of the central span of an anti-roll bar. This actuator performs its function by imposing a pre-load upon the two sections of the bar.

The body 2 and the rotor 1 are fixed to the two halves of the bar; the blades 3 divide the internal volume of the actuator into four different chambers.

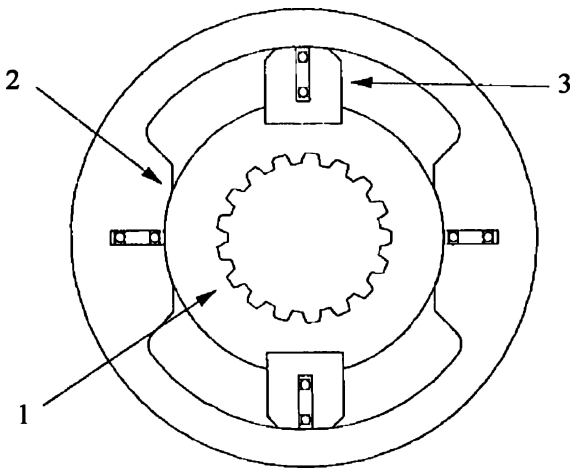


FIGURE 6.14. Hydraulic rotary actuator to control an anti-roll bar; this actuator is set in the middle of the central span.

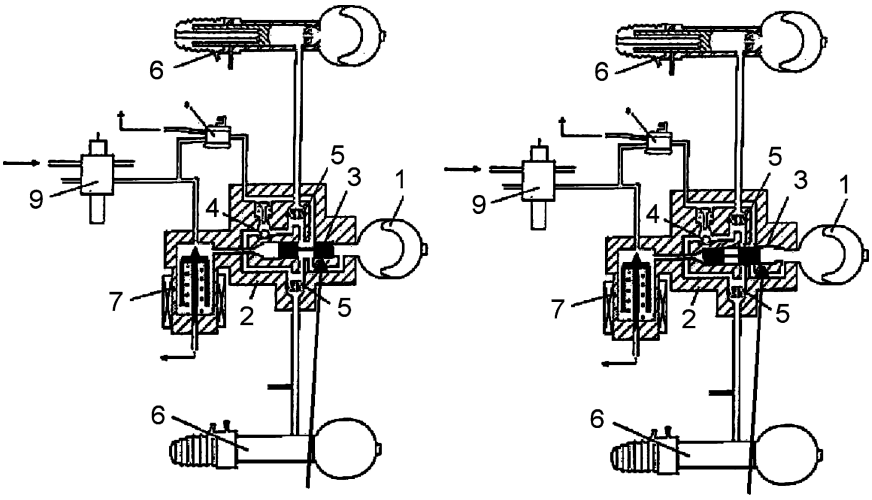


FIGURE 6.15. Scheme of hydraulic pneumatic suspensions, able to control vehicle trim and roll stiffness; in the A position (at left) the axle has a low roll stiffness, while in the B position (at right), the stiffness is higher (Citroën).

Through a four way solenoid valve it is possible to send oil under pressure (≅150 bar) to two opposed chambers and to exhaust it from the others or vice versa; two rotation directions are available, their angle proportional to the volume of displaced oil.

The actuator can also be made from a simple hydraulic cylinder working on one of the ends of the anti-roll bar, between the bar and the suspension.

Figure 6.15 shows the scheme of a recent hydro pneumatic suspension system by Citroën.

This system, here represented for one of the two axles, includes the trim and roll angle control functions.

The trim control function is obtained by charging or exhausting oil in the cylinders 6, communicating with the pneumatic suspension elements.

In addition to the four hydraulic units, one for each wheel, two supplementary elements 1 are provided, which increase roll flexibility and decrease damping, when put in parallel to the elements of the same axle.

The insertion of these supplementary elements is controlled by the solenoid valve 7, in the same figure. The solenoid valve moves a spool 3, communicating with the two primary elastic elements; this spool is represented at left in the open position, at right in the closed position. The body of the supplementary unit includes the roll control valve (4 in the same figure).

This is a simple check valve that, in a curve, interrupts communication to the wheel inside the curve, thus limiting the roll angle. The spool is locked in neutral position when the trim control is active.

Two different roll stiffness values can be obtained, in the A and in B positions; at the B position the flexibility can be further decreased by the valve 4.

A second classification criterion considers the type of control adopted, which can be active or passive.

In a passive control the system can have two configurations:

- Low roll stiffness (free anti-roll bar) for driving on a straight road
- High roll stiffness (pre-loaded anti-roll bar) for driving on a curve

Actuator control is performed by a simple on-off valve; the energy supplied to the system is quite low.

In an active control the system features two proportional valves, one for each axle; this system can stabilize the vehicle dynamic response, with a higher energy consumption.

6.3.4 Active suspensions

The active function is able to control the pitch and roll motion of the body; the scheme is shown in Fig. 6.16.

Active suspensions could, in principle, avoid the application of elastic elements. Such suspensions are inspired by a model of suspensions capable of retracting the wheels while climbing an obstacle, or extending them when descending.

This function is quite different from that of traditional shock absorbers; in such systems, in fact, the applied force always opposes the suspension motion.

This kind of suspension is usually classified in two categories:

- Slow active, or narrow band systems
- Full active, or broad band systems

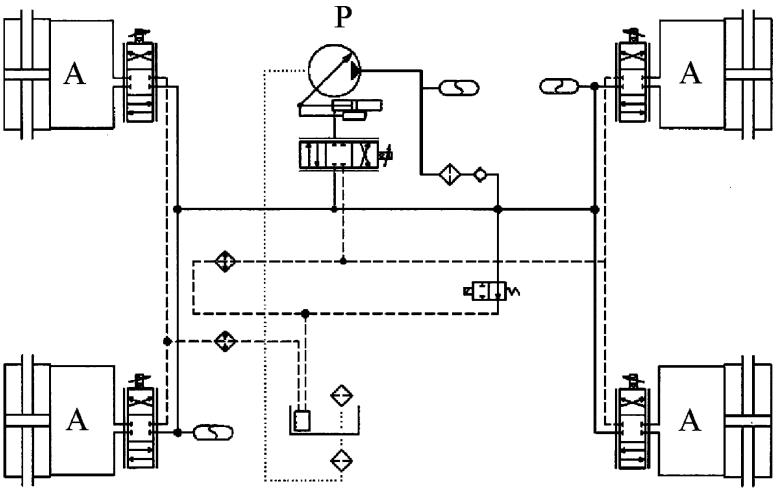


FIGURE 6.16. Hydraulic scheme of an active suspension; the suspension motion is controlled by four double effect actuators in communication with a pressure circuit, through solenoid valves; peak pressures are managed by pressure accumulators.

Slow active systems control 4 pressure regulating valves, one for each suspension.

These valves are able to control body motion (in the frequency range between 0–5 Hz), while higher frequency motions are managed passively by pneumatic springs, in parallel with the suspension hydraulic actuator.

The control strategy can follow the sky-hook damping theory for slow frequencies only; for higher frequencies, such as tire frequencies excited by small obstacles, they behave as passive suspensions with optimized damping coefficients.

In addition to sky-hook damping they can improve vehicle maneuverability and stability and perform anti-dive and anti-squat functions.

Full active systems are able to manage dynamic phenomena in the range between 0–25 Hz, having a flow regulating solenoid valve; in this way pulse maneuvers, such as driving on an obstacle, can also be managed.

A peculiar characteristic of these systems is the high demand for hydraulic power from the engine (8–10 kW) at high pressure (150–180 bar); large pressure accumulators are necessary to cope with peak demands.

The speed of the actuators must reach a minimum of 2–2.5 m/s.

7
CHASSIS STRUCTURES

Chassis structures are stressed by internal and external loads.

External loads come from the wheel-ground interface, moving through the suspension mechanism and its elastic elements, and from the aerodynamic field around the car body.

Internal loads are caused by the mass of the vehicle and payloads (such as passengers and baggage). Significant internal loads are produced by the reaction forces of the power train suspension.

Chassis structures can be *separated* from the body, as in industrial vehicles and some commercial and off-road vehicles, or they can be *integrated*, as found in unitized bodies. In this case, *auxiliary* structures are sometimes applied, to better distribute local loads to the body, supporting suspension mechanisms, engine or power train, transmission and final drive.

Chassis structures in these three cases are also called, respectively, *frames*, *underbodies* and *subframes*.

A typical example of a front subframe, added to support lower suspensions joints, some power train mounts and the steering box, is offered in Fig. 3.6.

7.1 UNDERBODY

The term unitized body derives from the fact that the body structure is designed to bear all vehicle loads directly; the other case – a separated body – does not mean that the body is only one of the external loads stressing the chassis structure, the frame. A separated body also contributes to supporting external loads.

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

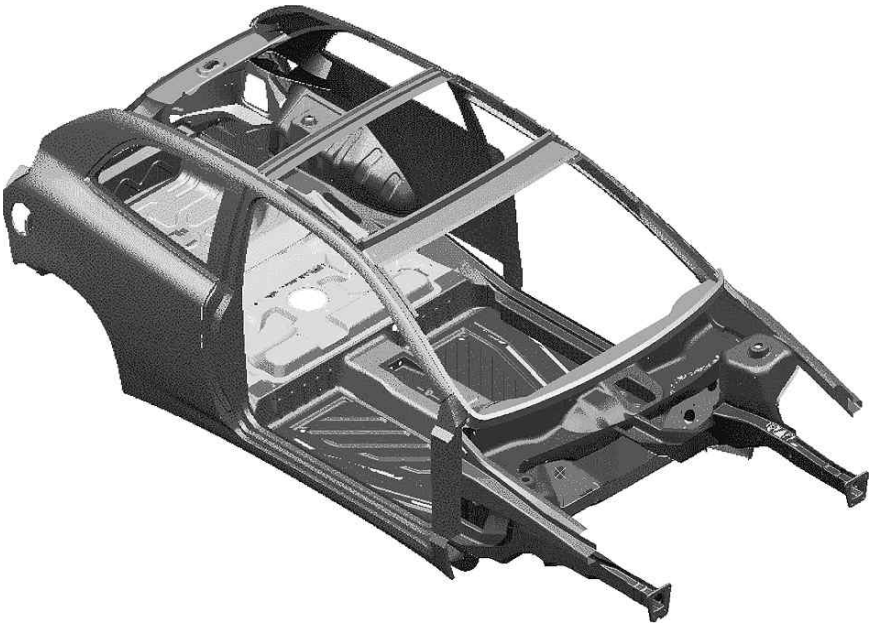


FIGURE 7.1. Unitized welded body shell for a two-box, two door sedan; the body shell is what remains of a unitized body, after removing interior trim, doors, hood and other elements not relevant to structural function (FIAT).

The adjectives 'separated' and 'unitized' are better referred to the technology used to assemble the body.

In unitized bodies, chassis structures cannot be disassembled after the production process, while in separated bodies they can be conceptually disassembled; separated body technology allows the vehicle manufacturer to supply a complete running chassis to a body manufacturer.

In this paragraph we will give a short description of the body of a two door mass production two-box sedan; we will make reference to the most widely used technology, a spot welded stamped steel sheet solution. The purpose of this description is to identify chassis structures in a unitized body.

Other solutions are present in the market, usually dedicated to niche cars, adopting different materials or joining techniques. Considering the objective of this book we will not discuss them.

Figure 7.1 shows the body shell; this term describes what remains of a car body after all parts without structural function, such as removable parts, interiors and glass, are removed.

Removable parts are side doors, rear or trunk door and hood. In many cars, front fenders and the front end can also be easily removed, to make replacing them easier, in the event of collision damages. These components have a limited structural function, in terms of body stiffness and resistance and are not considered to be part of the body shell.

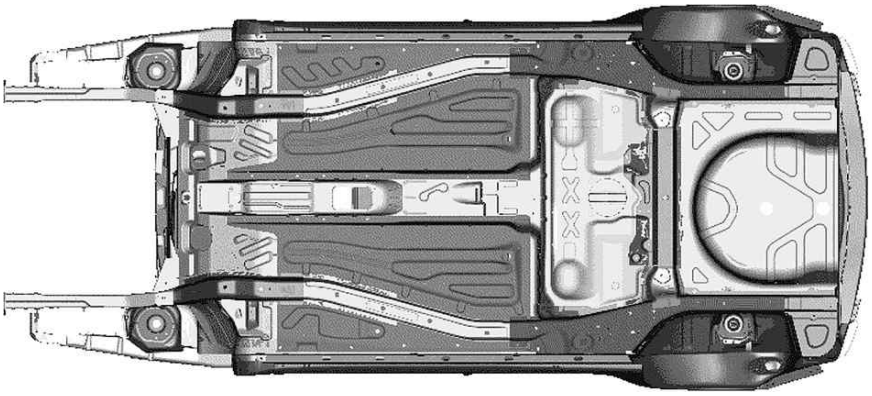


FIGURE 7.2. View of the underbody of the body shell of the previous figure, giving evidence of the integrated beams performing the function of the chassis structure.

The same can be said for glass, with the exception of the windshield, when bonded to the body; in this case its contribution to body stiffness is not negligible.

The body shell can be imagined as a space frame (all beams making up the frame do not lie on a single plane, as in conventional frames) with many thin panels, such as fire wall, roof, sides and floor contributing to the structure through their shear stiffness.

A view from the bottom side of the same body shell is shown in Fig. 7.2, making visible the integrated beams performing the function of the chassis structure; the longitudinal elements are still called side beams, and the cross beams cross the body shell. The layout of these beams is imposed by the requirement of offering stiff mounts to suspensions, power train, bumpers and other components of the chassis.

The shape of these beams also plays a major role in determining the behavior of the body (deformation, energy absorption and occupant protection) in the event of a crash.

The underbody has a functional analogy to the chassis of an old car or industrial vehicle. This subassembly is sometimes called a *platform*, and even if it is not able to perform any function without the rest of the body, it has some interesting technological and organizational features.

In fact, this part of the body shell can be conveniently shared between vehicles that may be quite different in appearance, such as two and four door sedans, station wagons, minivans, coupés, roadsters, etc., as major components of the chassis. Such usage offers beneficial effects on investments and costs.

With this objective in mind, modular underbodies are designed so that interchangeable side beams and cross beams of different length can be assembled together to generate platforms for the different wheelbases and tracks produced by common facilities.

A further technology feature is that this subassembly can have a production life longer than that of car models based upon it.

The organizational issue, consequent to the technological issue, is that the underbody can be designed and developed by a different team, together with other chassis components, as is now done by many car manufacturers.

The underbody, represented in the pictures above, is made of two primary longitudinal beams (which are the most important structural member of this assembly) as long as the entire vehicle; floor panels are welded onto them. The floor between the axles is trimmed by two additional side beams, part of the body sides.

Cross beams connecting these side beams feature a closed section that is obtained by welding an open stamped profile to the floor panel; by looking to these sections we can see how it is possible to obtain very stiff beams for the body space frame by joining thin stamped panels together.

The body space frame is therefore composed of a ladder structure for the underbody, and a top frame, below the roof, connected by vertical elements, the so called *pillars*; usually, those ahead of the front doors are called A pillars, while those behind the front doors are called B pillars. C pillars are those behind the rear doors.

To understand how this space frame is made we can analyze the cross sections shown in Figs. 7.3 and 7.4; numbers shown in cross sections refer to the same stamped elements.

Section A-A refers to the side beam of a frame surrounding the windshield integrated with the A pillar; this frame is composed of the side panel of the body 1 and two reinforcements 2 and 3.

Section B-B cuts the roof side beam, part of the frame of the rear door; the beam is formed from the side of the body 1 and an internal double frame 20 and 21; the roof panel 5 is also welded to this beam.

Section C-C refers to the lower part of the A pillar, where the front door hinges are assembled: the beam is again formed from the body side 1 and an internal side panel 15 with a reinforcement 10; the hinge is bolted to the plate 12; to this element the fire wall panel 16 is also connected by welding.

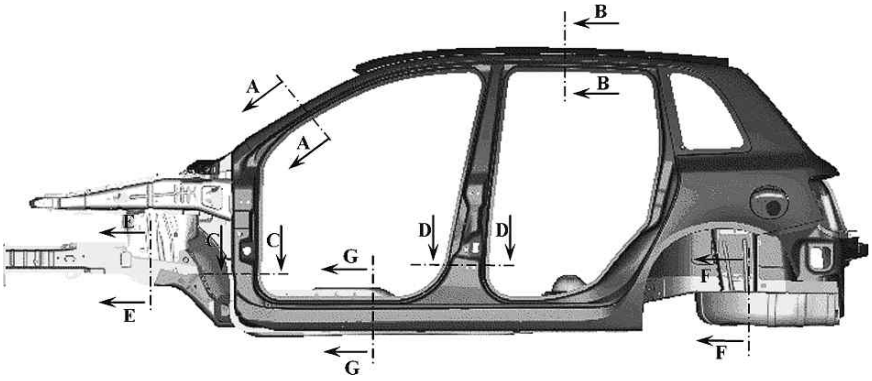


FIGURE 7.3. Side view of a four door body shell showing references to the cross sections presented in the following figure.

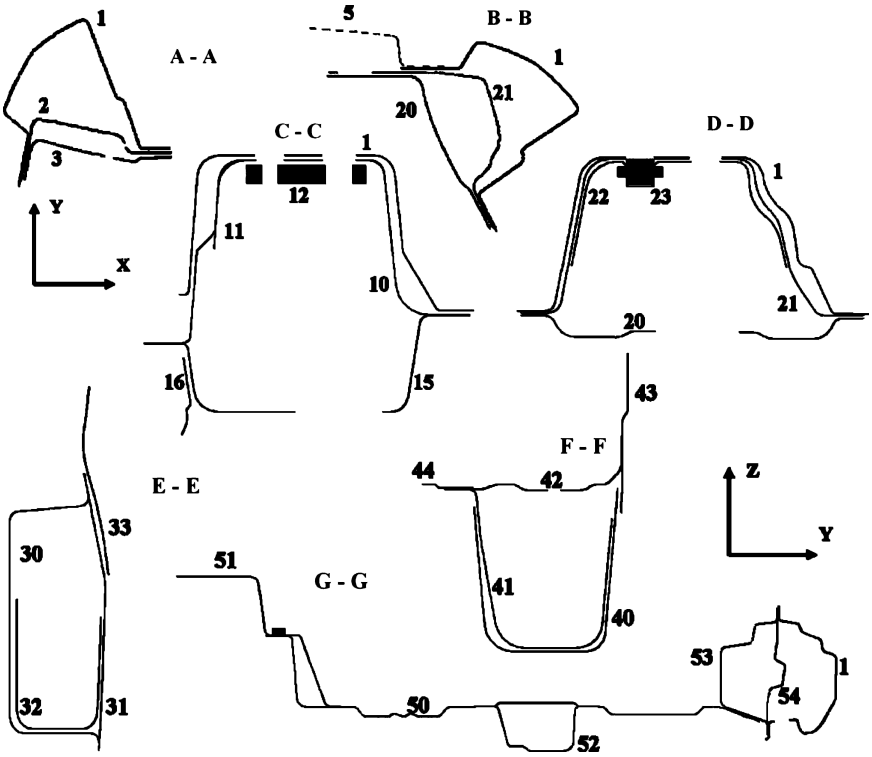


FIGURE 7.4. Section A-A: vertical member of the windshield frame, section B-B: roof side beam, section C-C: A pillar, section D-D: B pillar, section E-E: front section of main side beam, section F-F: rear section of main side beam; section G-G: left half of the floor.

Section D-D refers to the part of the B pillar where the rear door hinges and front seat safety belts are connected; the beam is again formed from the body side 1 and the internal side 20, with a local reinforcement 21; the hinge is bolted to the plate 23.

Section E-E is cut on the front of the main side beam (see also Fig. 7.3): Notice the inside panel 30, an outside panel 31, a reinforcement 32; to this beam is also welded the inside panel of the front fender 33, connecting to the upper pivot of the McPherson suspension.

Section F-F represents the rear part of the main side beam (see also Fig. 7.3) connecting with the rear twist beam suspension; this beam joins different panels, such as rear floor 44 and the wheel well 43.

Section G-G, finally, cuts across the floor and shows another section of the main side beam (element 52 and floor 50) and of the lateral side beam (elements 53, 54 and body side 1); a third side beam is shaped beside the tunnel made by the floor 50 and a reinforcement 51.

7.2 SUBFRAME

Subframes or auxiliary frames, on a unitized body structure, perform the following functions:

- They offer suspensions and power train mounts and distribute the consequent loads to the body area most suitable from a structural standpoint.
- They built up a secondary suspension, when mounted on the body with elastic elements, able to filter vibrations from the powertrain and wheels, at frequencies critical to acoustic comfort.
- They contribute to managing body deformations in the event of a crash.
- They offer an assembly support to many elements of the chassis, with benefits for work organization.
- Their reduced dimensions, as compared with the dimensions of the body, allow better control of tolerance of the suspension mounts, with benefits for their elasto-kinematic behavior.

Subframes are often applied to the front of the car and offer mounting points for suspension lower arms, anti-roll bar, steering box and part of the power train.

Figure 3.6 in the chapter dedicated to suspensions is an essential example of this kind of structure; it is made from a stubby beam that fixes the McPherson suspension arms. This subframe allows a stiff connection for the suspension in an area where the body structure is too distant to provide it.

The same figure shows the anti-roll bar mounts and Fig. 3.9, in the same chapter, shows the installation of the rack and pinion steering box. It is easy to understand how it is possible to guarantee a high precision for all assembly holes in this structure using a multiple drilling machine. The precision of their positions is important for toe angles and the alignment of the steering wheel spokes.

All these chassis elements are assembled on the subframe on a dedicated preparation line. This subassembly can be installed on the body in a short time, as compatible with the speed of the main assembly line.

This subframe is connected to the body through two large mounts, visible at its rear; these mounts are flexible rubber bushing, designed to filter power train and wheel vibrations. The remaining mounts, further toward the passenger compartment, are instead rigid (on the two vertical cantilevers).

The subframe is made of two stamped steel shells (see section A-A on Fig. 3.28), which are spot welded along a mounting flange. It is also possible to make a structure of this kind with an single piece aluminum casting, using almost the same shape; the choice is tied to a trade-off between weight and cost.

The power train (including engine, transmission and differential) is mounted on the underbody main side beams through elastic suspensions; only a reaction rod (for power train reaction torque) loads the lower part of the subframe.

FIGURE 7.5. Perimetrical subframe for large front wheel driven cars; the front of the car should be imagined on the left side of the picture (Lancia).

This kind of subframe is widely used for front wheel driven small and medium size cars.

In larger cars a different kind of subframe is used, surrounding the perimeter of the lower part of the engine compartment, as shown in Fig. 7.5.

Figure 3.39 in the chapter dedicated to suspensions offers a partial view of the suspension matched to this subframe; the car is again front wheel driven with transversal power train. A similar solution could also be applied to rear wheel driven cars or to front wheel driven cars with longitudinal power train.

With a perimetrical subframe of this kind it is possible to avoid any direct connection to the body b the lower suspension arms, anti-roll bar and steering box even when the distance between these mounts is larger than in the previous example; the power train too can avoid a direct mount on the body.

Figure 7.6 shows a view from the top of this kind of subframe (car front at right) with reference to the most important cross sections.

The joining points with the body shell are set at the four vertices of the frame; sections 5-5 and 7-7 illustrate the rubber bushing seats for an elastic suspension. The frame is a monolithic beam (see section 6-6), bent to a U shape, welded on a multiple element cross beam, cut at section 2-2.

The monolithic beam is made by hydroforming. This process, whose implementation is relatively recent, provides that a steel tube, with no welding, is bent to the final U shape. This semi-finished part is placed into a mold, reproducing inside the final external shape, and then pressed inside by a very high pressure fluid (about 1,000 bar); this pressure brings the steel to its yield point and imparts the final shape to the beam.

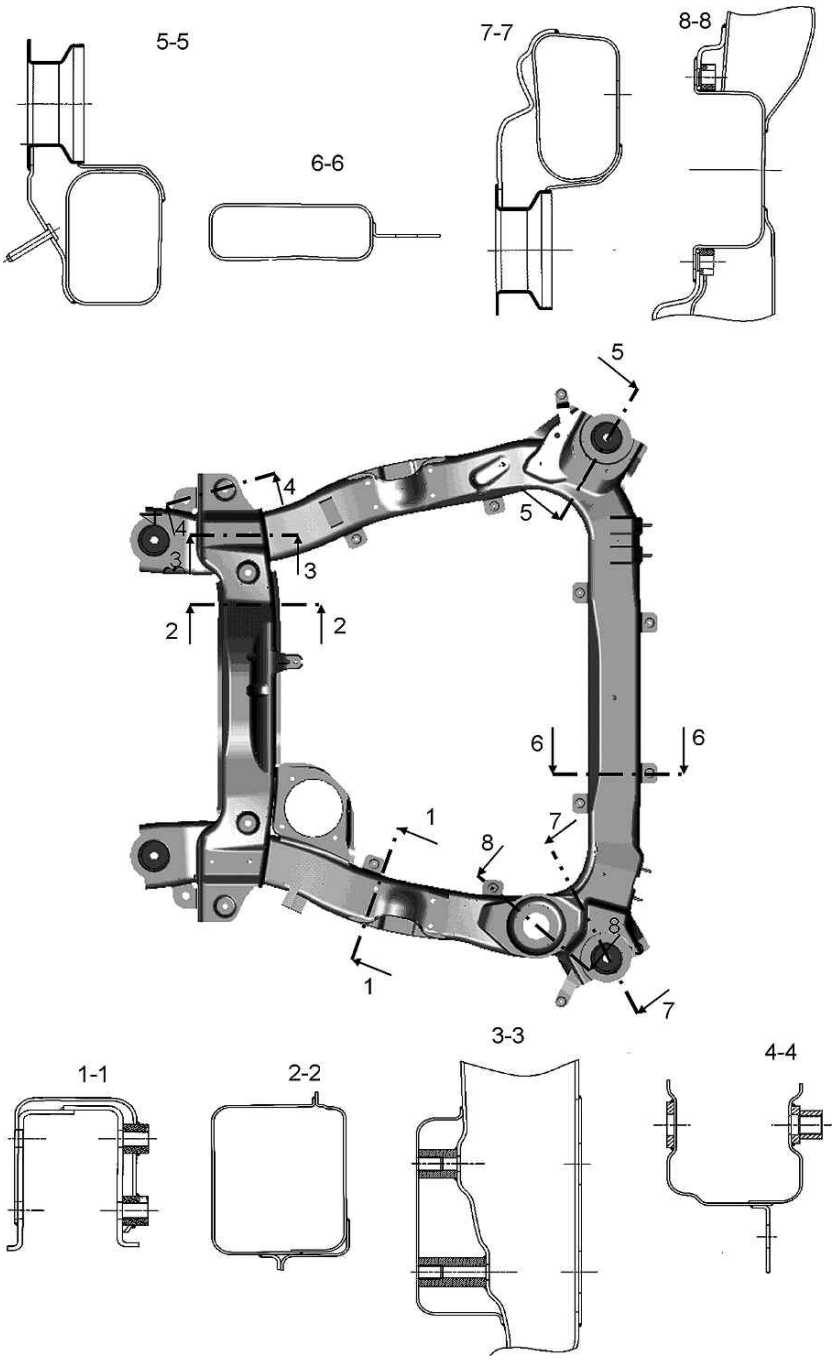


FIGURE 7.6. Perimetrical subframe showing its most important cross sections.

The advantage of this process is evident; tooling is simpler than conventional stamps and the beam cross section is not weakened by welding or made heavier by flanges or overlaps. In addition, the cross section can change along the beam axis.

On the other hand, the final shape has no aesthetic finishing (not important for this kind of application) and requires that the cross section perimeter be almost constant along the axis, to avoid excessive decrease in thickness or rips.

The closure beam of the subframe, cut by section 2-2, is made of two shells, welded together and to the hydroformed beam through arc welding.

The figure illustrates two power train mounts for large rubber bushings; one of them is cut by section 8-8. The power train mounts (four in total) are all found on the perimetrical subframe.

Where suspension arm rubber bushings are installed, the hydroformed beam is cut and reinforced; this feature is shown in cross sections 1-1 and 4-4. Openings allow the arm rotation axis to be positioned closer to the center of the car, with a more suitable elasto-kinematic performance.

Section 3-3 shows also one of the steering box mounts; all mounts are made with projection welded threaded bushings.

Many other holes allow a complete installation of power train accessories, harness and piping before final assembly of the subframe onto the car body.

Another example of a subframe is shown in Fig. 7.7; this example is applied to a rear suspension for a large front wheel driven car.

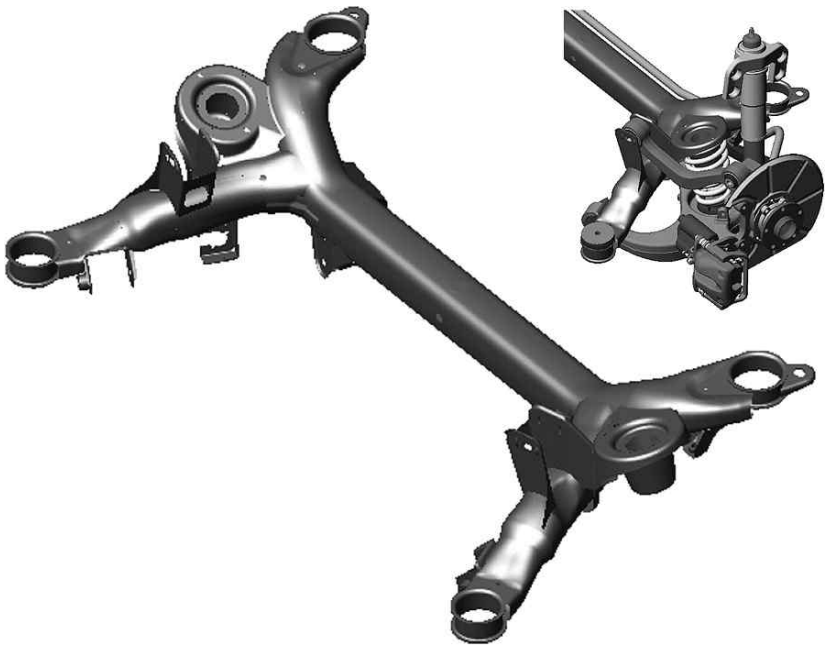


FIGURE 7.7. Subframe for a multilink rear suspension for a large front wheel driven car; at the upper right the complete assembled suspension is shown for reference (Lancia).

The car front should be imagined at the right.

The frame has in this case an H shape and features on the side beams of the H the mounts of suspension arms and spring and shock absorber seats.

The eyes at the tips of the longitudinal beams are used for rubber bushings installation for assembly with the car body.

In small cars the rear subframe can be avoided and the suspension mounts applied directly to the body shell.

In front wheel driven cars with suspended differential box, the subframe is modified to install the differential. Its dimension is increased longitudinally to offer good reaction to the torque that is applied by the differential in the y direction.

Figure 7.8 shows some cross sections of the rear subframe of the previous figure.

A simpler structure, as compared with the front subframe, can be obtained by bent tubes with local stamping, welded together; hydroforming is, in this case, unnecessary.

A U bent tube (the U end are upwards in the figure) with two short beams composes the final shape. Section 3-3 shows an overlap area for welding the two elements; a reinforcement diaphragm is also added. The same section shows the coil spring seat.

Sections 2-2 and 4-4 show the U shaped tube in two different points; section 1-1 shows the front mount of the lower suspension arm.

7.3 INDUSTRIAL VEHICLE FRAMES

Industrial vehicle frames continue to use the configuration developed for the first cars, a ladder or grillage structure composed of side and cross members.

As a matter of fact all considerations discussed in the chapter on design evolution still apply to industrial vehicles.

Given the enormous diversity of final applications of these vehicles, body manufacturers must remain dedicated specialists separate from vehicle manufacturers; many bodies (such as these of flat trucks, vans, dumpers, etc.) cannot be adapted for structural jobs, as described at the beginning of this chapter.

The cabin too, even if made with technologies similar to the car body is too limited in size to carry significant loads.

For all these reasons a real frame is needed to carry all loads and connect the chassis components. Some exceptions apply to buses and tank trailers, whose study is outside the scope of this book.

The frame is made by side members and cross members that should be attached to the former as stiffly as possible; the rigidity of this attachment is responsible for the most important part of the torsional stiffness of the frame, which is the only element capable of reacting to torques applied along the vehicle x axis.

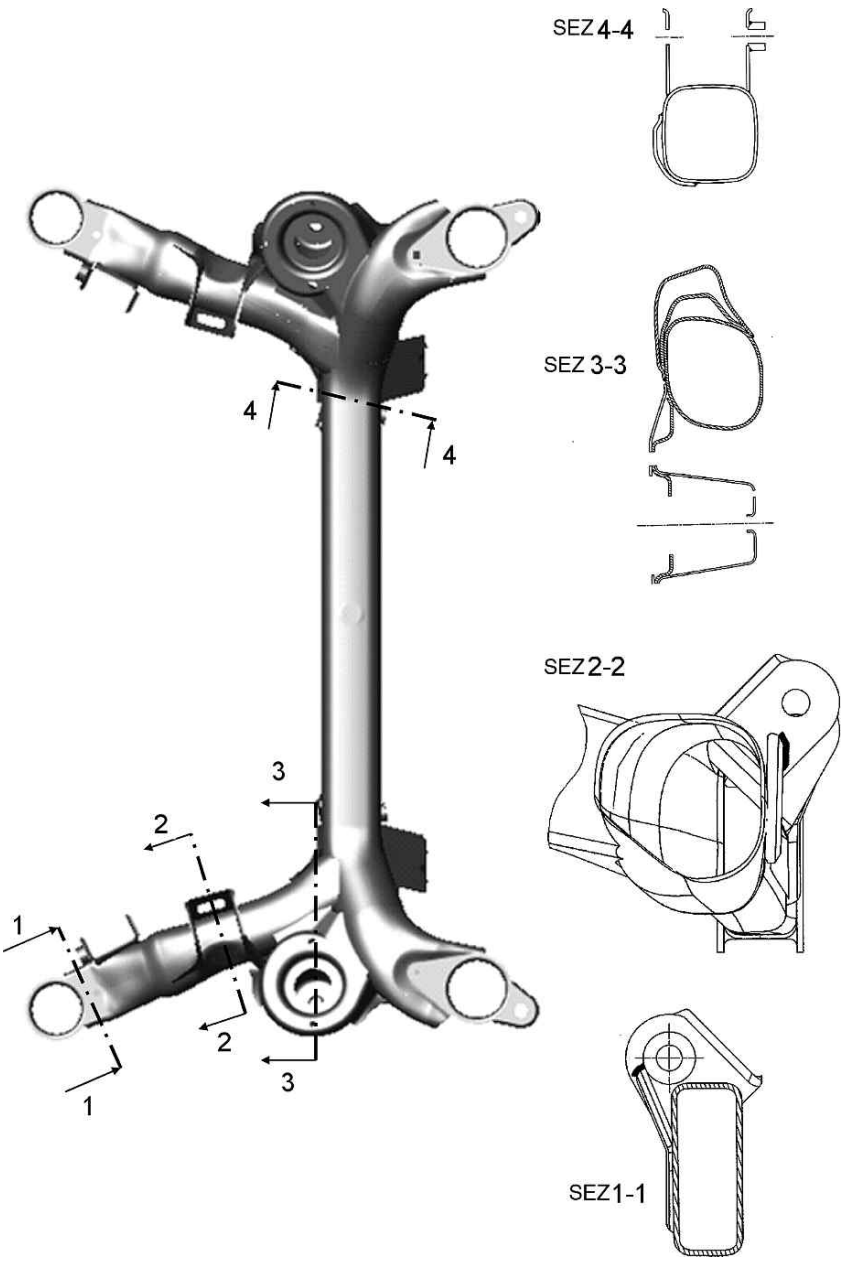


FIGURE 7.8. Relevant cross sections for the rear subframe of the previous figure.

Figure 7.9 shows an example of a chassis frame for a heavy duty truck and some design details useful for understanding how this structure works.

Side beams are shaped to different widths, when seen from above; the narrower part can accommodate the steering and twin wheels, while the larger is adapted for engine installation and the driver's cabin.

In addition, side beams are also tapered (as seen from above), to allow the suspension stroke of the rear wheels, caused by the large range of payloads.

The side beam cross section is shaped like a C, even if an open section is not the most suitable for high torsional stiffness. Open cross sections, widely applied, can be easily bent, beginning with high thickness, high resistance steel sheets. Where necessary, their cross section is reinforced with the application of additional larger C beams or plates.

Cross beams carry on the job of distributing local loads applied by suspensions and power train; an example of their layout can be seen in Fig. 3.66 and in others in the same paragraph, dedicated to industrial vehicle suspensions. In addition, cross beams contribute to the torsional stiffness of the frame.

As a matter of fact, if the two side beams were completely disconnected, the moment of inertia of the frame would be double that of a single section. Because this section is built up with an open profile, characterized by rectangular elements of thickness s, much smaller than their length h, we have approximately:

$$I_x = \frac{\eta}{3} \Sigma h_i s_i^3 \ , \tag{7.1}$$

where h_i and s_i are the dimensions of the three rectangles building up the side beam cross section; the parameter, in this case, is $\eta \approx 1,12$.

Cross beams ideally constrain the side beams to remain parallel. The torsional deformation of the side beams is limited by the bending stiffness of the cross beams, increasing the torsional stiffness of the assembly.

Cross beams can be made in a simple C profile, with their ends bent and without platbands, inserted in the inside of the side beams (as in Fig. 7.9, detail a), or they can be reinforced locally with C profile cuts, welded at their ends, to improve the connection with the side beams (as in Fig. 7.9, detail b).

Further local reinforcement can be achieved by shaping the ends of the cross beams (as in Fig. 7.9, detail c) or applying two Ω profiles shaped for the purpose and welded together (as in Fig. 7.9, detail d).

When it is necessary to reduce the dimensions of the cross beams, a tubular cross section can also be applied, welded to flanges at their ends (as in Fig. 7.9, detail e).

It should be noticed that all joints with side and cross beams are made with rivets or bolts if they are to be disassembled for repairs (Fig. 7.9, details 1 and 2); this choice is demanded by the large dimension of the frame. Arc welding, with its introduction of high heat, would induce excessive deformation and consequent residual stress that would be difficult to eliminate with heat treatment.

Arc welding (Fig. 7.9, detail 3) is limited to the preparation of small components or cross beams.

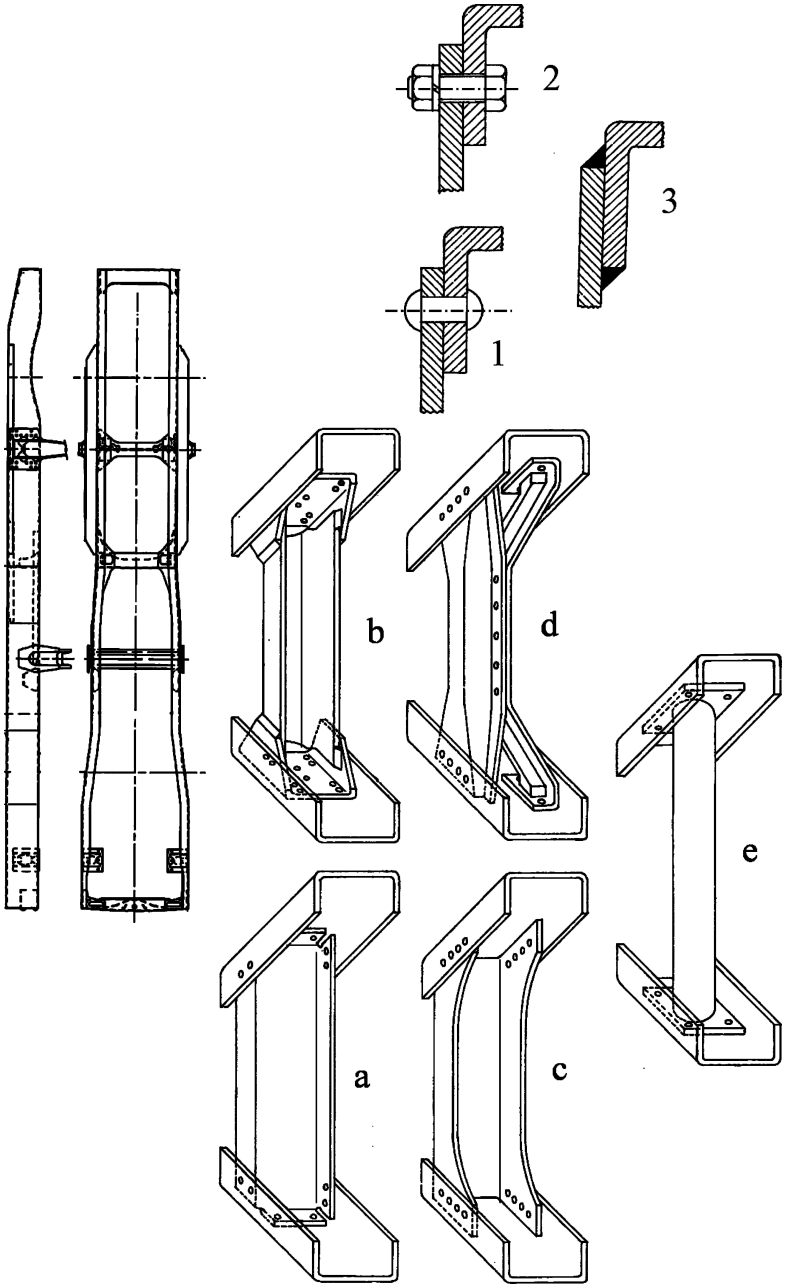


FIGURE 7.9. Side and top view of a chassis frame for a heavy duty truck (at left). Details a, b, c, d and e show different types of joints used to connect cross with side beams. Details 1, 2 and 3 show different joining technologies, addressed to exchanging high intensity shears in the side beam plane.

7.4 STRUCTURAL TASKS

The tasks of the chassis structure is to bear forces and payloads and contain consequent deformations.

The issue of deformations must be approached not only from a static but also a dynamic point of view, considering vibration amplitudes and the noise they produce in the passenger compartment.

Some considerations on this topic are discussed in the second volume.

7.4.1 External loads

External loads applied throughout the life of the chassis structure can be classified under two different categories:

- Instantaneous overloads
- Fatigue loads

An example of the first category can be supplied by imagining what happens when driving over large pits, hitting the curbs of side walks, braking suddenly on high friction ground or starting up in low gear.

It is common practice to consider these loads as static loads, by introducing multiplication coefficients to their steady state values. For loads bound to friction between wheel and ground, prudential friction coefficients $\mu_{y,p} = \mu_{x,p} = 1.2 - 1.3$ are taken into account; as far as shocks are concerned, accelerations $5 - 7$ times higher than acceleration of gravity are considered.

Emblematic fatigue loads can be imagined by driving on bumpy roads with almost full loads; typical roads are those paved with cobble stones or asymmetric ridges.

The expected life of a car structure is at least 200,000 km of average use; each manufacturer has designed and built test courses severe enough to allow the vehicle average life to be simulated in only $50 - 100,000$ km; the completion of a mission like this requires at least several months of driving, with drivers organized in multiple shifts.

Loads recorded on these courses, which are empirically correlated to car life, can be further synthesized by eliminating loads under the structure fatigue limit; load histories are elaborated that allow average life to be reproduced in less time.

Static loads and mass properties

Vehicle weight is always present. The reaction to this load comes from the contact point between tires and ground, through the elastic and structural elements of the suspension. Suspensions, according to their architecture, distribute reaction loads in all directions, not necessarily just vertical, through their connections with the chassis structure.

TABLE 7.1. Axle loads F_{z1}, F_{z2} and center of gravity position a, measured from the front axle, divided by the wheel base l as a function of load conditions, on a medium front wheel driven car. M is the sprung mass.

Load condition	F_{z1} (N)	F_{z2} (N)	F_z (N)	a/l %	M (Kg)
Curb weight	6,800	4,770	11,570	41.2	1,030
1 pass. (driver)	7,150	5,120	12,270	41.7	1,100
2 pass.	7,500	5,470	12,970	42.2	1,170
3 pass.	7,630	6,040	13,670	44.2	1,240
3 pass. + 30 kg	7,610	6,360	13,970	45.5	1,270
5 pass. + 50 kg	7,860	7,710	15,570	49.5	1,430

We have seen, for example, in the McPherson suspension, that the vertical reaction applied to the wheel, the lower arm and the shock absorber assembly generates a bending torque that is equaled by a lateral force applied to the chassis structure.

We should remember that unsprung masses (wheels, tires, struts, brakes and part of the mass of the suspension arms, springs and shock absorbers) rest on the ground directly, without intervening vehicle structure.

The distribution of vehicle weight and reaction forces is a function of the passengers on board and transported baggage; this set of parameters is called weight condition.

Table 7.1 summarizes axle loads and longitudinal center of gravity positions for different load conditions in a medium size front wheel driven car, assuming a mass of 70 kg for each passenger, including the driver. In this kind of car, passengers sitting in front have their center of gravity almost at mid-wheel base. Their weight is shared at 50% by the two axles. Rear passengers are seated at about 80% of the wheelbase, while the baggage center of gravity is usually set at about 105% of the wheel base.

Each added mass shifts the position of the center of gravity, according to this formula:

$$x_G = \frac{m_1 x_{G1} + m_2 x_{G2}}{m_1 + m_2} \,, \tag{7.2}$$

where x_{G1} is the initial position of the center of gravity of a vehicle of mass m_1 on the x axis and x_{G2} is the position of the added mass of value m_2.

The table also shows an estimate of the sprung mass, made by car body and interiors, power train, passengers and their baggage.

Load conditions also affect the moment of inertia of the sprung mass, useful for predicting the vehicle dynamic behavior.

Considering the standard reference system x, y, z with its origin in the center of gravity of the sprung mass, the vehicle system is characterized by an inertia ellipsoid, whose symmetry axes are generally not coincident with the reference axes.

The vehicle inertia tensor is therefore characterized by three inertia moments J_{xx}, J_{yy}, J_{zz} and by three centrifugal moments J_{xy}, J_{xz}, J_{yz} which often neglected, at least at curb weight condition.

The mass contribution of passengers and baggage modifies the moments of inertia according to the following formula:

$$J_{ii} = J_{ii,1} + J_{ii,2} + m_1 d_1^2 + m_2 d_2^2 \,, \tag{7.3}$$

where subscripts 1 and 2 refer to the original reference mass and to the added mass, J_{ii} is the moment of inertia of the added body, according to the i direction, through its center of gravity, m is the added mass and d is the distance between the center of gravity of the added mass and that of the reference condition.

With a similar formula the centrifugal moments can also be calculated.

The experimental determination of the center of gravity position can be made by weighing. The ISO 10392 standard provides that, having measured sprung masses m_1, m_2, m_3, m_4 at the wheel positions (1 is front left, 2 is front right, 3 is rear left and 4 is rear right), the center of gravity position can be measured by the following formula:

$$x_G = \frac{m_3 + m_4}{m_1 + m_2 + m_3 + m_4} l \,, \tag{7.4}$$

where x_G is the distance of the center from the front axle and l is the wheelbase.

The transversal position of the center of gravity, with reference to the vehicle mid-plane (positive to the right) is a function of the front and rear tracks t_1 and t_2:

$$y_G = \frac{t_1(m_1 - m_2) + t_2(m_3 - m_4)}{2(m_1 + m_2 + m_3 + m_4)} \,. \tag{7.5}$$

Center of gravity elevation can be obtained by raising an axle and measuring the masses on the axles remaining on the ground m'. If the rear axle is raised, we can use the following formula:

$$z_G = \frac{l(m'_1 + m'_2 - m_1 - m_2)}{(m_1 + m_2 + m_4 + m_4)\tan\vartheta} + R_{l,i} \,, \tag{7.6}$$

where ϑ is the angle of raising, $R_{l,i}$ is the rolling radius of the axle remaining on the ground; an analogous formula applies when raising the front axle.

To measure moments of inertia, see the procedure described in the second volume.

Driving loads

Because of driver maneuvers, the chassis structure is subject to loads that change in time and are to be added to the static load. Control forces on steering wheel, accelerator and brake pedals cause additional loads on the contact points between wheels and ground, because of the implied side and longitudinal slip.

These loads are transferred to the chassis structure by the suspension.

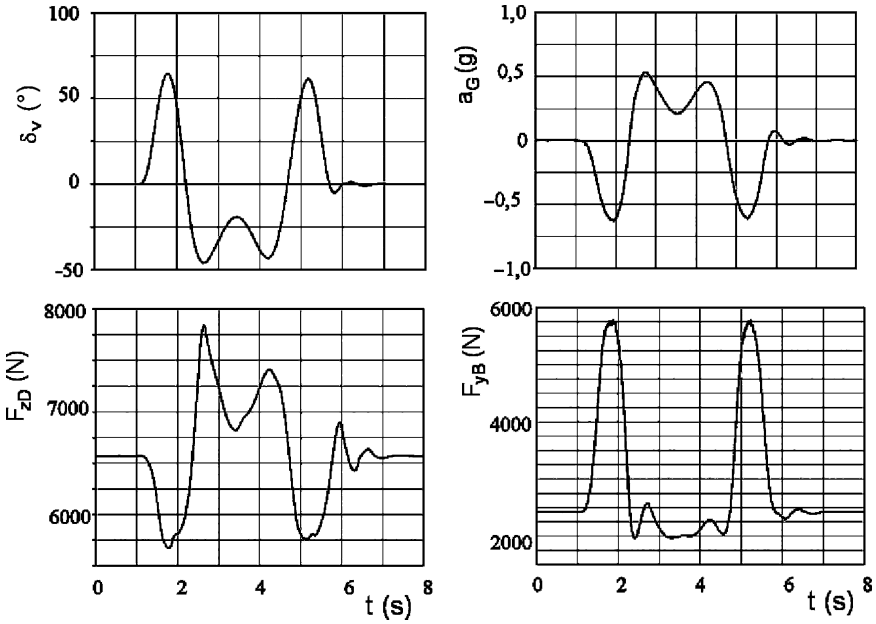


FIGURE 7.10. Records of steering wheel angle (above left), transversal acceleration at the center of gravity (above right), vertical force on suspension upper pivot (below left) and transversal force on the lower suspension arm (below right) taken from an ISO overtaking test.

The final result is longitudinal and transversal accelerations that affect vehicle motion.

In modern production cars a transversal acceleration of 1 g is reached while turning at limit conditions, while about 0.5 is reached in longitudinal accelerations.

Inertia forces cause, in addition, pitch and roll angles of the sprung mass and transfer loads from one wheel to the other.

Figure 7.10 shows sprung mass transversal acceleration and some of the front suspension loads, calculated for a car during an ISO overtaking maneuver. The simulation has been made by a multi-body mathematical model of the vehicle, including a complete description of the quadrilateral suspensions.

Loads are often calculated from a simplified model (where suspensions are functionally described by their elasto-kinematic curves), subsequently applying then these loads to multi-body or FEM models of the vehicle suspension.

Loads due to uneven ground

Vehicle structure is also subject to dynamic loads coming from uneven ground. Under these forces, vehicle subsystems react dynamically and load their connections to the chassis structure.

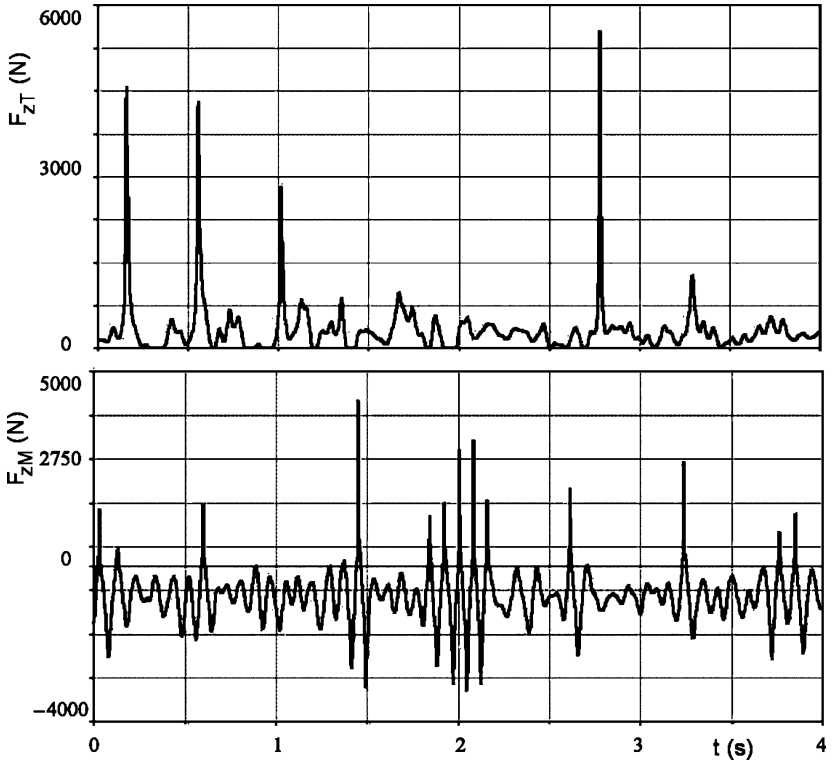


FIGURE 7.11. Simulation of loads on the suspension stop of a rear axle and on an engine mount while driving over uneven ground.

Figure 7.11 shows results of a simulation of a car driving on an uneven road, used for predicting fatigue life.

The calculation is made by using a multi-body model of the entire vehicle, including all mechanical details of suspensions and engine mounts. A high level of forces is evident, either through the suspension joints or engine mounts. The former are influenced by suspension stops, which for large displacements make suspension characteristics stiffer. Similar phenomena occur for the latter.

7.4.2 Internal loads

Vertical accelerations generate, as we have seen, inertial forces on the power train mounts to the chassis. Torque applied to the driving wheels also reacts on the suspension, gearbox and differential. In front wheel driven cars all reaction torques are applied to power train suspensions.

Other internal loads are caused by reciprocating masses in the internal combustion engines.

The engine mechanism is made of a crank rod system that applies a reciprocating motion to the piston; to represent forces applied to the engine crankcase, we normally separate the reciprocating mass of the rod into two parts concentrated into the crank and piston pins (see the second volume). The masses of the two parts, together with part of the crank and the piston, define the so-called reciprocating mass m_a and rotating mass m_r.

Having defined as:

Ω the rotary engine speed,

r the crank radius,

the rotation of the crankshaft will induce on each crank an unbalanced centrifugal force:

$$F_c = m_r \Omega^2 r \ . \tag{7.7}$$

By neglecting engine block flexibility, centrifugal forces will be added to those of other cylinders, often generating a totally balanced system.

If we call:

ϑ the rotation angle of the crank,

l the distance between centers of the crank pins,

α the aspect ratio of the mechanism r/l,

each reciprocating mass will apply to the engine block a force:

$$F_a = - \, m_a \Omega^2 r (\cos \vartheta + \alpha \cos 2\vartheta) \ . \tag{7.8}$$

This formula approximates the piston acceleration with a Fourier series of two members only; forces applied to different cylinders must also be added, taking into account the different orientation of different cranks.

These are not the only moving masses; other suspended masses, such as the steering wheel or the exhaust pipe, can also generate non-negligible internal forces.

7.4.3 Stiffness

Structural stiffness plays a fundamental role in the driving and vibratory behavior of a vehicle.

In addition, it is important to limit deformations, because significant loads may affect vehicle operation, by preventing doors from opening or closing, for instance, or altering the kinematic behavior of suspensions.

Flexural stiffness K_f is defined as the ratio between a load applied to the middle of the wheelbase and the deflection of the same point; reaching acceptable values is generally not difficult, if other structural requirements are satisfied, except in the case of very long vehicles.

Torsional stiffness K_t is instead the ratio between a roll torque applied to the wheel hubs of the front axle and the consequent rotation, when the rear axle hubs are fixed to the reference system. In this ideal scenario, elastic primary and secondary elements of each suspension are replaced by rigid elements of equal geometry.

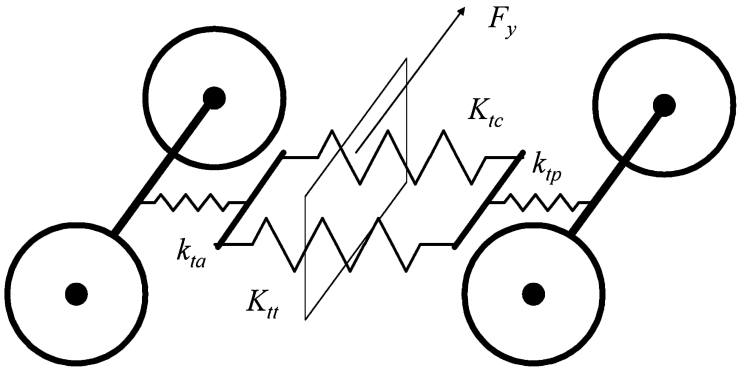


FIGURE 7.12. Elementary scheme for modelling the torsional deformation of a vehicle: k_{ta} and k_{tp} represent the torsional stiffness of the two axle suspensions, while K_{tc} and K_{tt} represent torsional stiffness of body and chassis frame in a hypothetical vehicle with separated frame.

To justify the importance of torsional stiffness on the dynamic behavior of a car, let us consider the simple model in Fig. 7.12.

The vehicle is modelled with four element with torsional deformations, respectively:

The body shell, with torsional stiffness	K_{tc}
The frame, with torsional stiffness	K_{tt}
The front axle suspension, with torsional stiffness	k_{ta}
The rear axle suspension, with torsional stiffness	k_{tp}

In this simplified scheme, we assume that the axles are fixed to the frame at non-deformable sections. The body shell is fixed to the frame as well; in other words there is no displacement other than that induced by the considered elements. We assume in addition that the sprung mass is concentrated in its center of gravity and that its position is not affected by deformations.

If the vehicle is driven on a curve, if F_y is the centrifugal force, a roll moment will be applied to the vehicle as:

$$M_x = F_y h_G \ , \tag{7.9}$$

where h_G is the elevation of the center of gravity from the ground.

The roll torque must be equal to the elastic reaction of the axles and of two structural elements of the vehicle, given by:

$$M_x = \varphi \left[(\frac{2}{K_{tc} + K_{tt}} + \frac{1}{k_{ta}}) + (\frac{2}{K_{tc} + K_{tt}} + \frac{1}{k_{tp}}) \right] \ , \tag{7.10}$$

where φ is the roll angle.

This formula considers the configuration of the deformable elements in series or in parallel: those in parallel, such as body shell and frame, add their rigidity, while those in series, such as suspensions and vehicle structures, add their flexibility.

We have seen the importance of load transfer in determining the cornering stiffness of an axle and we can argue, even if we cannot yet explain, the influence of the roll stiffness of the two axles in determining the understeering behavior of the vehicle.

As a matter of fact, the axle that will absorb the most important part of the load transfer will be subject to larger side slip angles at the same value of cornering forces; in our example the ratio between load transfers is given by:

$$R_t = \frac{(\frac{2}{K_{tc}+K_{tt}} + \frac{1}{k_{ta}})}{(\frac{2}{K_{tc}+K_{tt}} + \frac{1}{k_{tp}})} \ . \tag{7.11}$$

This ratio is controlled by the suspension roll stiffness of both axles (ratios k_{ta} and k_{tp}) only if the frame and body shell rigidities are much larger, so as to make the first members of the numerator and of the denominator defining R_t negligible.

In other words if the vehicle structure is not sufficiently stiff the understeering behavior will be determined by this last, instead of by suspensions and anti-roll bars.

If we now suppose our car to cross an asymmetric obstacle, so as to impose a roll angle φ on the front axle, a torque M_x will be applied to the vehicle so as to satisfy the following condition:

$$M_x(\frac{1}{k_{ta}} + \frac{1}{k_{tp}} + \frac{1}{K_{tc}+K_{tt}}) = \varphi \ . \tag{7.12}$$

The torsion angle applied to the structure is given by the term in brackets; because the frame and body shell must rotate at the same angle for congruency, the ratio R_c between the torques affecting the two elements is given by:

$$R_c = \frac{K_{tc}}{K_{tt}} \ . \tag{7.13}$$

The stiffer element will carry on the larger portion of the torque; stiffness and resistance must therefore be proportional.

A very stiff body shell should also be highly resistant; this explains why wooden bodies on steel chassis frames have been made with flexible joints in their skeleton. And we can understand why a roof can crack because of an overly flexible floor in a unitized body.

We can assume that a reasonable bending stiffness target should be set within the interval $700 - 1,000$ daN/mm, while a reasonable torsional stiffness target is between $70,000 - 150,000$ daNm/rad.

7.5 STRUCTURAL DESIGN

It is not our intention to supply in this section comprehensive knowledge about vehicle structural design; as we have seen that in modern cars, the underbody and body shell are completely integrated. The technical disciplines necessary to implement the design process are the work of the body development teams.

But we do want to supply some highlights on methods that can be easily applied in concept development. They can be used to predict the consequences on the vehicle structure of the application of a suspension or a subframe.

The objective of this approach is to understand in advance if an assigned chassis component is compatible with the rest of the vehicle.

To predict the structural performance of a vehicle it is important to verify the feasibility of the project targets.

The numeric analysis methods available today, along with the always increasing processing capacity of computers are perfectly adequate to this purpose, but the need for detailed mathematical models of the structure makes their application difficult during the preliminary design phase, because much of the information they need will be only available later. Vehicle structure completely dependent upon visible body shape, which is frequently modified, because body style is under development or many competing shapes are undergoing parallel development. In addition, aerodynamic performance optimization, also performed in parallel in this phase, introduces frequent shape changes.

During the preliminary design of a vehicle, a capacity for synthesis and quick decision is essential.

This implies the identification of a set of critical performances to be predicted and taken under control within the project targets, beginning with superficial, incomplete and frequently modified information.

For this limited purpose we will introduce the *structural surface* method and the *beam model* method, which are in part interchangeable in this kind of application; the second is particularly useful for separated frames or industrial vehicle frames.

7.5.1 Structural surface method

The body shell structure can be idealized, for the study of its performance during preliminary design, with a system composed of beams shaping a spatial grid and of closure panels; the contribution of these panels to the global structural behavior is quite important.

The structural surface concept introduces many assumptions to simplify the model.

A structural surface is an elementary flat panel that because of its limited thickness can accept loads contained in its mean plane only; loads directed to other directions cannot be withstood because of its high flexibility.

The simplest structural surface is given by the rectangular panel shown in Fig. 7.13, defined by two sides of dimensions a and b and thickness s, which is assumed to be negligible with respect to the other dimensions. For such a panel the inertia moments of the cross sections are:

$$J_x = a\frac{s^3}{12} \ , \quad J_y = s\frac{b^3}{12} \ , \quad J_z = b\frac{s^3}{12} \ . \tag{7.14}$$

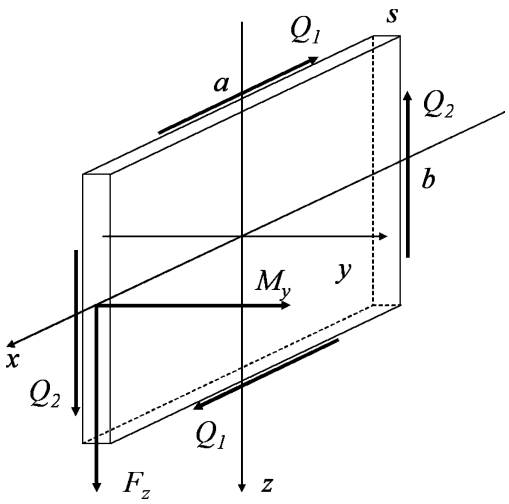


FIGURE 7.13. A simple structural surface composed by a flat rectangular panel.

Because of our assumption, J_x and J_z are much smaller than J_y and the structural surface so defined will be able to withstand only bending moments along the y axis and forces parallel to the sides.

For example, the application of a force F_z to one side of this panel can be balanced by two shear forces Q_1 and Q_2 applied to the two sides, so as to satisfy the equation:

$$\begin{aligned} Q_1 b - Q_2 a &= 0 \ , \\ F_z - Q_2 &= 0 \ . \end{aligned} \tag{7.15}$$

But in consideration of the limited thickness, these shear forces can be applied only if there are other constraints, able to limit the shape instability of this thin surface on the compressed side (the lower one in our example). This constraint can be another panel, laying on a perpendicular plane, welded to the first, or a beam probably shaped by a closed section with the near panel.

The entire body shell can be modelled with a set of plane structural surfaces, which approximate its curved surface. The contribution of the beams can be neglected, retaining only their role in avoiding wall instability.

Panels can be made in different shapes, but are always characterized by their ability to react to forces contained solely in their plane

Figure 7.14 shows some useful examples for the following applications. Figure 7.15 shows an example of a schematic model of a three volume sedan.

A car body structure can be modelled for a simplified structural analysis as a combination of structural surfaces; now we can look at the consequences of these assumptions.

Let us consider a roll moment applied to the body shell of Fig. 7.15 at the front axle position; this moment can be balanced by two shear forces acting on panels simulating the interior walls of the front wheel wells.

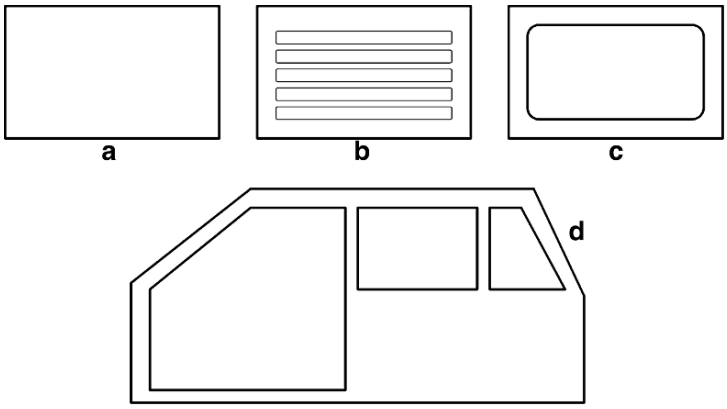


FIGURE 7.14. Examples of panels for a structural surface model: a rectangular panel, b panel with reinforcement grooves, c panel with lighting opening, d door frame panel.

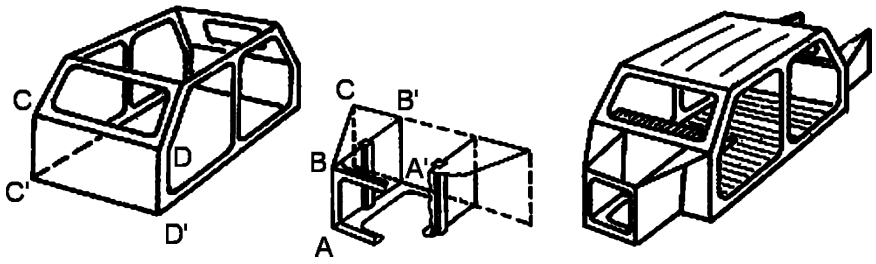


FIGURE 7.15. The many kinds of panel, needed for building a model of a hatchback four doors sedan.

For example, the panel AA'BB' receives a force contained in its mean plane and directed upwards; as for the panel in Fig. 7.13, it is possible to calculate the shear forces that must be applied to its sides by the adjoining panels.

The shear reaction on the side of this panel will load the near CC'D'D panel and, from this, the panels near it. It should be noticed that the neighboring triangular panel BB'C cannot offer any contribution, because it is perpendicular to the direction of force, as are roof and floor panels.

A similar logical process can be applied to all forces that the front suspensions apply to the body structure through the front panel AA'B'B.

If we consider a lateral force applied to the suspension, because of a bend or the inclination of their arms, we must conclude that the panel AA'B'B is unable to react properly; in this case, the triangular panel BB'C will do the job.

Similar considerations can be applied to a concentrated load applied to the car floor, such as a seat mount or a rear suspension arm. The structure will be able to react only if at the application point of the load a panel with its mean plane directed parallel to the force is applied.

This explains why floors must also use vertical panels (as side walls of the side beams and lateral side beams) where loads are applied.

Although resulting from rough approximations, this rule must be followed meticulously in order to design a sound structure.

The structural surface method has the following advantages:

- Allowing the evaluation of the behavior with simple calculations in two steps
- Enabling design modifications of a single surface without necessarily affecting neighboring surfaces

The first step consists in defining loads on the sides of the structural surfaces. These loads can be kept constant, as a first approximation, even if their surface will be modified; this can speed up evaluations when a design is changed frequently.

The second step consists in determining the shear stress on each surface, where the previously defined loads act.

The many subassemblies such as vehicle sides, door frames, roof and floor can be studied separately as plane problems.

7.5.2 Beam model method

Beams are mathematical objects that allow a simplified description, in terms of stress and strain, of the vehicle structure under the action of external loads and constraints.

Using structural analysis these mathematical models are defined according to the finite elements method (FEM). If we assume limited displacements, we can accept the usual relationship between forces and displacements:

$$\mathbf{Ku} = \mathbf{f} \ , \tag{7.16}$$

where:

$\mathbf{K}$ is the structural stiffness matrix;
$\mathbf{u}$ is the displacement vector;
$\mathbf{f}$ is the vector of forces applied to the structure.

In preliminary design the mostly widely applied numeric models are beam FEM models. In fact, the beam is the simplest element: A segment joining two nodes with six degrees of freedom, to which the characteristic features (section area and moment of inertia) of a body shell closed section can be attributed. Examples of beams of this kind can be designed from the cross sections in Fig. 7.4.

The beam element formulation and the De Saint Venant theory are coincident.

By using elements of this type the structure can be described as a space frame made of beams. Their elastic and geometric characteristics are solely defined by their cross sections.

This approach is similar, if more complex, to that used to describe plane frames, such as those used by industrial vehicles or, as subframes, by cars.

A beam model of a car structure is composed of:

- Beam elements describing the structure: These join the nodes in the space as elements with 12 degrees of freedom (6 for each node) and are characterized by appropriate sections and bending and torsional moments of inertia; panel contributions are neglected or included in the beam.
- Nodes joining the beams that can be simulated by rigid or flexible joints.

During preliminary design of a new car, the data available to a structural analyst are minimal: Usually only the wheel base, tracks, suspensions and body shell components that are shared with already existing cars (carry-over); in some cases the entire underbody is carried over. In addition, a style development center has developed shapes for external volumes, openings (doors, hood and hatch) and glass; this information is usually included in a mathematical model implemented in a computer aided styling system (CAS).

A beam model can be synthesized beginning with:

- A CAS model on whose surface and openings a structural space frame is adapted.
- Existing models of carried-over elements.
- Structural cross section; if missing this can be scaled from already existing cars of the same segment or from a database prepared for the purpose; this information can again be retrieved from drawings such as those in Fig. 7.4.

The steps of this activity are the following:

- Repositioning existing carried-over structural models on the new style model
- Adapting carried-over parts to the new dimensions
- Identifying node positions of the new structure
- Deriving beam center lines from style surfaces
- Describing areas and moments of inertia of beam cross sections
- Identifying relevant loads and constraints

The result of this process is described in Fig. 7.16 for a small car body. The versatility of this approach is easily understood.

- A change in style can be introduced into the model by laying down beams below the new surface.

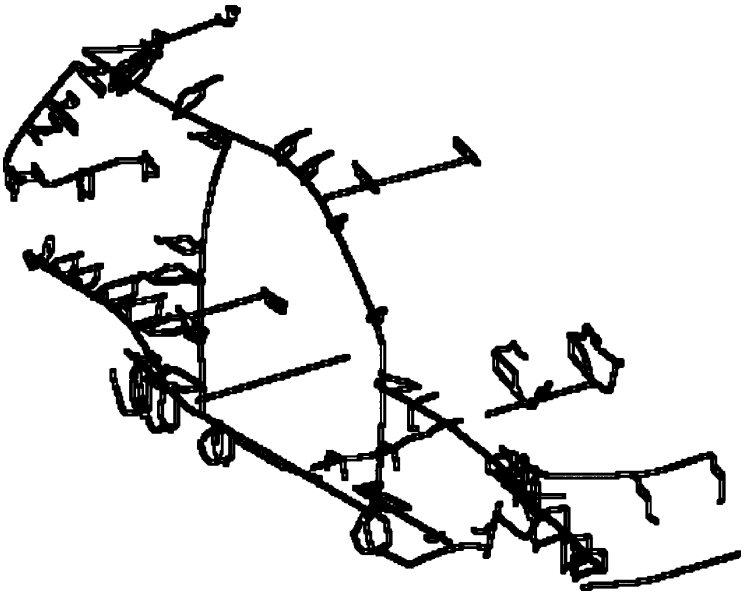


FIGURE 7.16. Final scheme of a beam model; the shape of the beams is determined by placing them below the styled exterior shape and assigning to them compatible cross sections, scaled from existing data bases.

- Cross sections can be easily modified by their description tables.

- The stiffness of a node can also easily be modified by their description tables.

A synthesis beam model also allows us to draw tentative conclusions about how different parts of the body contribute to the final result and let us perform a first weight optimization. By scaling up the dimensions of each section one at a time it is, in fact, possible to affect system performance and to discover iteratively the most effective dimensions.

The beam model method assumes an approach opposite to the structural surface method, by neglecting the role of panels.

The two methods should not be considered as in competition with each other to solve the same problem.

The first, for unitized bodies only, can be applied for concept development involving the general architecture of the body. It is simple and allows many different alternatives to be considered in a short time.

The second method offers a better approximation and can be applied after the first to optimize an already defined structure. Unlike the first, the second method is sensitive to exterior shape modifications as well. It is also better suited to frames and space frames.

7.6 STRUCTURAL TESTING

The test benches used for structure fatigue evaluation are similar to those introduced in our section on suspensions; for economic reasons and considering that real suspensions are the simplest way to apply loads to structures accurately the body shell, frames, suspensions and power train mounts are usually tested together.

The most important test, specific to chassis structure and the body shell, is the overall evaluation of torsional stiffness.

Beyond what we have already said about the inconvenience of insufficient structural stiffness, we would like to point out that the analysis of local deformations through bending and torsion tests is useful for discovering discontinuities in the structural function of the vehicle.

In fact, areas where deformation discontinuities appear evidently represent potential rupture points during on the road operation, because of the cyclic deformations caused by road obstacles.

We now concentrate on torsional and bending stiffness testing.

The torsion test simulates what happens when a vehicle climbs a curb with one wheel only, while the bending test examines what happens through the application of payloads.

Both tests are performed on the same bench, which is reconfigured according to the purpose of the test.

The test bench (shown in Fig. 7.17 for torsion and in Fig. 7.18 for bending tests) is made from a foundation block characterized by a stiffness well above the values to be measured.

On this block additional structures are installed to fix the body or the frames properly; other mobile structures are used to apply loads.

The interiors (dashboard, seats and side panels), complete power train (engine, transmissions, exhaust pipe, etc.), bumpers and wheels are removed from a unitized body. Suspensions are maintained, but springs and shock absorbers are exchanged for rigid spacers, in order to freeze the geometry under investigation.

It is common practice to verify the contribution of removable parts (side doors, hatch door, trunk lid, hood, removable fenders) and of structural glass (bonded windshield, bonded rear window glass) to the total stiffness by testing the structure with and without their contribution; this clarifies whether joining techniques are suitable for the purpose.

Constraints are similar for the two tests: wheel hubs are welded or fixed by brakes and are used to connect the structure to the cross beams used for interfacing the block and load application frames.

In torsion tests, the rear axle is linked to the block while the front axle is linked to the torque application structure; both links avoid longitudinal and transversal motion.

In bending tests both axles are fixed to the block, but one is capable of longitudinal motion.

Loads are applied in different ways in these two tests.

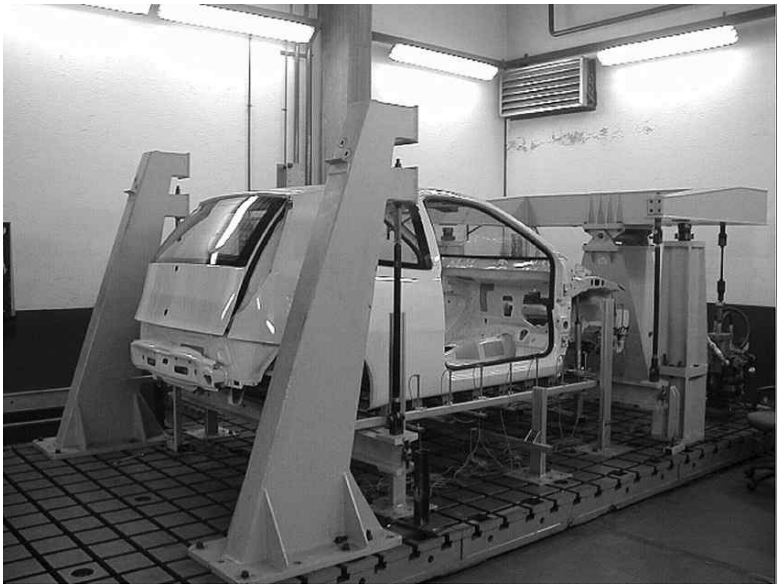


FIGURE 7.17. Universal bench prepared for torsion test; the articulated frame in the front is used to apply a roll torque.

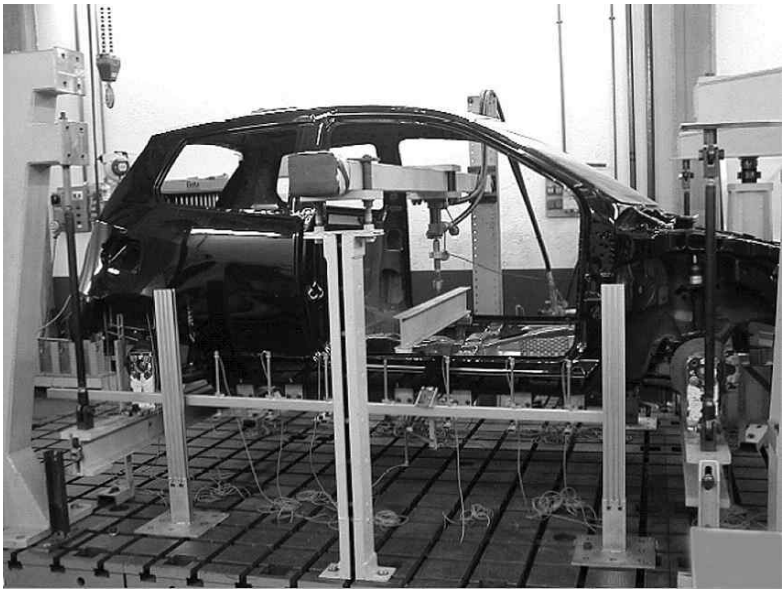


FIGURE 7.18. Universal bench prepared for bending test; the beam in the center is used to apply a local vertical load.

In torsion tests, a pure torque is applied to the front axle through an articulated frame.

In bending tests a vertical load is applied by an auxiliary beam to the center of the structure.

In both cases precautions are taken to avoid the influence of local deformations where loads are applied.

Displacements are measured by linear transducers on the vertical axis.

In torsion tests a pair of transducers is placed at the upper pivot of each suspension.

In bending tests one transducer is placed at the middle of the car and others are used to check local deformations.

Torsional stiffness is conventionally measured by the following formula:

$$K_t = \frac{M_x}{(\varphi_a - \varphi_p)} = \frac{M_x}{\arctan \frac{z_{da}+z_{sa}}{t_a} + \arctan \frac{z_{dp}+z_{sp}}{t_p}} , \tag{7.17}$$

where:

K_t is the torsional stiffness,
M_x is the applied torque,
φ is the roll angle at the two axles,
z is the vertical position variation of each wheel,
t is the track.

Subscripts refer to:

a front axle,
p rear axle,
d right side,
s left side.

To compare different cars, torsional and bending stiffness are often divided by wheelbase.

During this kind of test, local deformations are also recorded.

Many linear sensors are applied along the primary side beams and lateral side beams, in order to measure vertical displacements. In case of torsion, they are used to calculate the local rotation angle of the structure.

These displacements or rotations are used to draw diagrams depicting the function of the coordinate x of the body, to supply an indication of the stiffness distribution. Local variations of the slope of this curve, which can be detected through the first derivative diagram, highlight local critical areas (such as local instability of panels where sensors are applied) or structural discontinuities caused by inaccurate welding or design mistakes.

During torsion tests it is also important to measure the deformations of frames where doors and glass are applied: The difference between the diagonal lengths of the opening is usually taken as an indicator to be compared with project specifications.

Total stiffness changes when parts are added to the body shell; usual references include the body shell, body shell with fixed glasses, body shell with doors.

TABLE 7.2. Torsional stiffness K_t (daNm/rad), bending stiffness K_f (daN/mm) and masses M (kg) of body shells of two volume small cars of different years of commercial launch; the second subscript shows body conditions (n body shell, v with fixed glass, c complete). The contribution of glass to torsional stiffness can be noticed.

M. Y.	$K_{t,n}$	$K_{t,v}$	$K_{t,c}$	$K_{f,n}$	$K_{f,v}$	$K_{f,c}$	M_n	M_c
1990	34,200	40,400	47,800	430	445	475	152	207
1995	42,800	57,900	68,300	580	600	620	193	271
2000	57,300	70,700	79,600	630	640	670	190	264

Table 7.2 shows some values of torsional and bending stiffness measured for bodies of different model years. They are small two volumes cars.

Fixed glass is mounted on the body shell to examine its contribution; this comprises the windshield and rear windows that cannot be opened; their contribution is due to the bonding joint, which limits window frame deformation on the body. This contribution is negligible for bending stiffness, because of the limited deformation of the glass frames.

The body weight increase is partly due to increased dimensions, and partly to more severe passive safety regulations.

Many other functions could be added, given the almost unlimited performance of microprocessors.

INTRODUCTION TO PART II

Any kind of motor vehicle needs a transmission, in order to convert the rotational speed and the torque of its engine into useful power.

Under the term transmission are usually included a start-up device (either clutch or torque converter), variable ratio gearbox, power take-off, used to operate equipment external to the vehicle, and the final drive and differential, which move the wheel; in other words, the term includes the entire kinematic chain connecting wheels to engine.

The function of the transmission is to adapt the available torque from the engine to the needs of the vehicle, which are imposed by the nature of the road, the will of the driver and environmental requirements. The transmission is particularly relevant in determining certain vehicle functions, such as dynamic performance, fuel consumption, emission, drivability and, last but not least, reliability.

Automotive transmissions (we include under this definition cars, industrial vehicles and buses) are characterized, when compared to industrial transmissions, by a higher level of stress. The specific mass (the ratio between the transmission mass and the processed power) of an automotive transmission falls in a range between 0.5 and 1 kg/kW; in an industrial transmission these values are almost doubled. In addition, the number of ratios available to an automotive transmission is definitely higher than that available to an industrial transmission.

Transmission technology can be assumed to be mature; but we do not suggest that there will be no further evolution, particularly in automatic transmissions. As a matter of fact, significant growth of this kind of transmission is expected in the European market. The particular needs of this market have justified the development of a new generation of automated manual transmissions that should not affect vehicle performance negatively and should leave drivers the pleasure of driving manually as they choose.

The continuing development of electronics, both in terms of performance and cost, will have a significant impact on this kind of transmission and will cause rapid development.

We should also not forget that the possible growth of the hybrid vehicle market will stimulate development of particularly sophisticated automatic transmissions.

A last point relevant to transmissions is that their production and development functions are leaving OEMs to become the business of specialists.

This is true for both manual and automatic transmissions. It is particularly important for those who will devote their career to vehicle or component design to develop sound systematic knowledge of this kind of component.

This trend is confirmed by the growing importance in Europe of Getrag, ZF, GM, Fiat Powertrain Technologies and Graziano; the same could be said in the United States for Dana, Eaton, New Venture Gear and Allison, and in Japan for Aisin and JATCO.

Significant growth of automatic transmissions is expected in Europe.

The most widely used type of automatic transmission is the power-shift, with four or more transmission ratios. Many manufacturers have decided to adopt in their cars a continuously variable transmission, with steel belt and variable gouge pulley. Despite some disadvantages for mechanical efficiency, these show superior vehicle performance because of the availability of an infinite number of transmission ratios.

Power-shift transmissions are abandoning pure hydraulic control, in favour of more sophisticated electronic controls that take advantage of the contemporary development of engine control systems, including throttle automation.

On the other hand, many applications with more than five transmission ratios have emerged and the lock-up function of the torque converter is extended to nearly all gears.

Six gear manual transmissions are starting to appear in the market, while the five ratio configuration is standard on all applications; many of these transmissions have received an electro-hydraulic or electro-mechanical actuation system, with the aim of offering costumers a similar or superior performance, in comparison with the conventional automatic transmission, without affecting the traditional advantages of lower cost, high efficiency, and the pleasure of driving.

The functions that can be obtained by this kind of transmission include:

- Clutch automation
- Servo assistance in selecting and shifting gears, with positive effects on comfort and actuation speed
- Availability of more ergonomic sequential shifting commands, which can be included on the steering wheel
- Total automation of the choice of transmission ratio and shifting sequence.

Transmission architectures are determined by the kind of traction that is adopted on the vehicle. On the other hand, the choice of traction has a strong impact on vehicle performance, in terms of handling, ride comfort, safety and interior space organization.

The engine can be installed in the front or back of the vehicle; in the first case the traction can be on the front axle, on the rear axle or on both; in the second case the traction can be on the rear or on both axles. Rear wheel driven cars, with longitudinal engine in the front, are usually considered cars with a conventional or traditional layout.

Figure 1 shows all known configurations available for front wheel drive.

Sketch a shows the longitudinal layout, with powertrain on the driving axle; it also shows how some advantage can be obtained for the maximum steering angle of the wheel, with some disadvantage for the front overhang. In this case every transmission ratio can be obtained with a single stage of gears; the rotation axis of the gearbox output shaft must be turned 90 deg with a pair of bevel gears.

Sketch b shows the same previous layout, where the front overhang is shortened by installing the engine over the front axle; in this case an additional shaft and some modification of the oil sump are needed.

Sketch c shows the widely used solution of the transversal power train with engine and transmission in line. Here a single stage transmission is necessary; the output shaft is connected to the differential with a pair of spur gears. Vehicles can be more compact because the front overhang can be shorter; on the other hand the minimum turning radius is negatively affected. The solution is optimum if the engine is not too large.

Finally, sketch d shows a not widely applied solution, where engine and gearbox are transversal and parallel; the gearbox can be in the front behind or beneath the engine. Advantages and disadvantages are mixed, as compared with

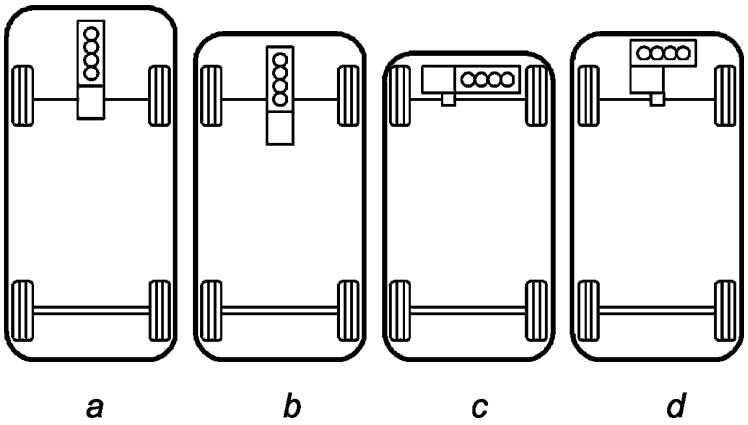


FIGURE 1. Different types of transmission architectures, available for front wheel driven cars; the impact on external dimensions is deliberately exaggerated.

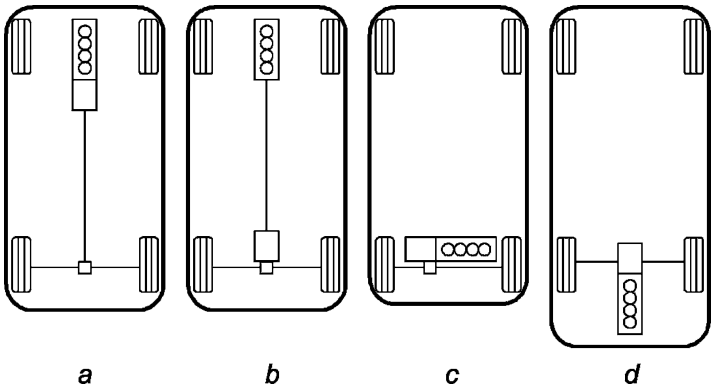


FIGURE 2. Different types of transmission, suitable for rear drives, with front or rear engine.

the previous case; an additional double stage transmission or a chain transmission is necessary to connect the engine shaft with the gearbox input shaft.

In all these cases the subsystem including engine and transmission is called the *power train* or *transaxle*.

Figure 2 shows two possible architectures applied to conventional vehicles; type a is the most widely used, while type b is applied primarily to sports cars, where a more uniform weight distribution is requested. In these cases the gearboxes will be of the double stage type (also known as the countershaft type), where input and output shafts are in line. A bevel gear final drive is also necessary. The front overhang can be reduced; the roominess of the interior is negatively affected by the transmission shaft under the floor.

Also the rear engine lay-out, which is seldom applied, shows two possible architectures, shown in the same figure, with transverse power train arrangement (c) and with longitudinal arrangement (d).

This kind of architecture is now only applied to sports cars, where weight distribution and yaw inertia are more important than interior roominess; the same considerations made in solution a and c of the previous figure apply to solutions c and d.

Architectures applied to commercial vehicles (transportation of people or goods with GVW of less than 4 t[1]) derive from car architecture d of the first figure and a in the second.

Figure 3 shows the architectures that are known for four wheel drive vehicles; configurations a, b and d can be derived from the corresponding lay-outs of the single axle vehicle. These are primarily applied to road vehicles or sport utility vehicles with permanent four wheel drive; solutions a and d are easily designed, because the transfer box can be integrated in the power train. Architecture

[1]This limit value is not coincident with the legislative value (3.5 t), but it is our guess about what is technically feasible for vehicles designed with car technology.

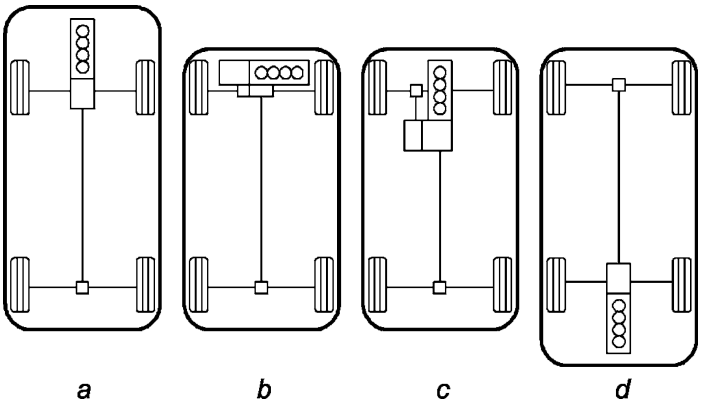


FIGURE 3. Transmission architectures applied to four wheel drive; solution c is particularly suitable for off-road vehicles, because an additional range-change unit can be easily integrated into the transfer box.

c is applied to vehicles specialized for off-road missions; the transfer gearbox necessary to move the other axle can easily integrate a range-change unit and differential lock.

Considering transmission configurations used in industrial vehicles (transportation of people or goods with GVW of more than 4 t), we must discriminate between trucks and buses, also taking into account that a low price market segment exists, where a bus is strictly derived from a truck.

Architectural configurations for two axle industrial vehicles are similar to those described for conventional drives. The longitudinal power train is installed in the front in a position depending on the type of driver cabin adopted. In the same way, the four wheel drive configuration is similar to that explained in Fig. 3c. In extremely heavy duty applications a second rear driving axle is moved through the first.

Buses, initially derived from trucks, have recently received a dedicated architecture finalized to assure easy passenger access to the vehicle. This access depends upon the height of the door threshold; the desired value for this dimension also differentiates the architecture of urban and suburban buses.

In the first case the easy access must be obtained with no compromise for all doors, and the corridor floor must include no steps or significant ramps, because of the necessity of transporting standing passengers.

Figure 4 shows two different examples; a represents a typical lay-out for urban buses, while b represents a typical lay-out for suburban buses. In the first example, the engine and transmission are installed in the back, under the last row of seats. The corridor floor is low and is conditioned solely by the shape of the axles, which is usually chosen for this purpose; the slope of the corridor is limited, and it is also possible to have doors in the back of the vehicle.

Suburban buses or long range buses, represented in b, show no real need for a lower door threshold; on the contrary, they usually have the baggage

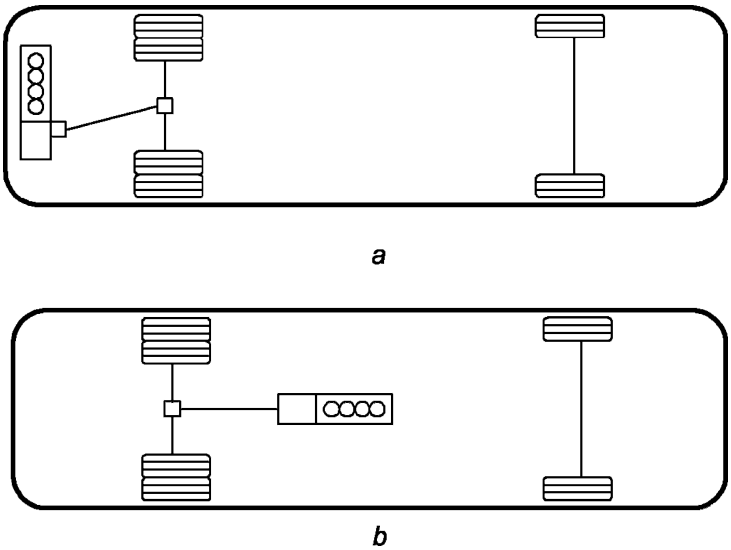


FIGURE 4. Transmission architectures designed for urban buses (high) and suburban buses (low); the engine shift-back lowers the vehicle floor behind the front door.

compartment under the floor. Passengers seldom stand; the most suitable lay-out has a longitudinal central engine and rear wheel drive.

Independent of transmission configuration, vehicle transmissions include the following components, which are named starting with the engine.

Start-up device

Start-up devices are usually made with a friction clutch, with pedal operation, in some case assisted by an external source of energy (electric, pneumatic or hydraulic);in most automatic gearboxes, start-up devices comprise a hydraulic torque converter with lock-up clutch.

Gearbox

The gearbox is the transmission component that changes transmission ratio; it can be made with conventional or epicycloidal gears. Gear shifts can be performed manually (with only the force of the driver), semi-automatically (with the help of auxiliary energy) or automatically. We can find in the gearbox the following mechanisms, in addition to gears.

- *Gear shifting mechanisms*: These are classified as internal or external, depending upon whether or not they are contained in the gearbox. The function of gear shifting mechanisms is conditioned by the motion of the power train relative to the vehicle, determined by engine suspensions; these motions should not affect gearshift quality. Gear shifting mechanisms are particularly important in manual transmissions, because they have a large influence on the driver's feeling of precision and positive response.

- *Shifting sleeves* and *synchronizers*: These are addressed to engaging and disengaging the pair of wheels necessary to obtain the desired transmission ratio.

Range change units

These can be considered as additional gearboxes with a limited number of speeds that are inserted on the transmission line after or before the main gearbox; they increase the total number of speeds by a number greater then the added gear pairs; under this category we also include the so-called *splitters*.

Final drive and differential

Final drive and differential can be integrated in the gearbox; they further reduce the rotational speed of the wheel. The differential theoretically divides the torque for the two wheels into equal parts and can be free, self-locking or manually-locking.

Transfer boxes

Transfer boxes are gearboxes that divide the torque for the axles of an all wheel drive or for vehicles with more than one driving axle.

Transmissions and joints

Transmissions and joints connect the several rotating sections of the transmission; in conventional drives they connect the gearbox with the differential. In independent suspensions, they connect the wheel with the differential.

The second part of this volume, dedicated to automotive transmissions, begins with a chapter dedicated to the historical evolution of transmission design.

The principal types of manual transmission applied to cars and industrial vehicles will be described in the following chapter. Schemes and drawings will be used to support this discussion.

Particular attention will be addressed to internal and external shifting mechanisms and to the criteria for their design and evaluation.

The primary start-up devices will be explained and the design and specification rules discussed. For the torque converter only, the methodology for calculating vehicle performance and consumption will be introduced.

In terms of differentials, in addition to design criteria, a short study on the influence of controlled and uncontrolled units on vehicle dynamic behavior will be included.

Various automatic transmissions in use on cars and industrial vehicles will be introduced; the primary control strategies implemented in their control systems will be also discussed.

The book concludes with some complementary ideas on transmission component design and on methods for car and bench testing.

8
HISTORICAL EVOLUTION

A gearbox and clutch, or alternative start-up devices, are essential for obtaining a driving torque suitable to traction from a reciprocating internal combustion engine.

In fact, this category of engines is characterized by a useful torque almost constant with rotational speed and sometimes increasing with it; in an ideal case, on the contrary, the driving torque should decrease with vehicle speed.

We can say that having decided to install an engine of a given power, considering the desired dynamic performance, this power should be made available at any vehicle speed; the traction force should therefore be inversely proportional to vehicle speed.

This choice would also guarantee vehicle speed stability, because a driving force decreasing with speed would balance a resistance force increasing with speed.

In addition, the internal combustion engine is unable to supply a positive torque if its rotational speed is less than a minimum value, whose amount is determined, as a first approximation, by the torque period of the cycle and the inertia of the crank mechanism. The vehicle, however, must be able to start-up when stopped and should be able to exploit, in this condition, the maximum driving torque.

For these reasons it is necessary to have a mechanical system available able to adapt the transmission ratio between engine and wheels to the needs of the vehicle and the deficiencies of the engine. This transmission ratio should be extremely high, ideally infinite, at vehicle start-up.

Finally, the transmission has the purpose of transmitting the motion of the engine crankshaft to the wheel hubs; as we know, wheels and engine have a

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

relative position not defined with precision, because of the motion of the suspension and the steering of the wheels.

In the first cars, the gearbox was confused with a device for speed adjustment; from this misinterpretation was also born the name of this device, almost equivalent in many European languages (for instance, *changement de vitesse*, in French). This term was used instead of torque converter or torque adaptor, more suitable to the real role of this mechanism.

These considerations, taken today for granted, were not so evident to the designers of the first cars, which were equipped with a steam engine. Among many disadvantages for vehicular application, such engines have the advantage of being able to supply a notable torque when the crank is stopped, quite suitable for vehicle application.

This situation is exemplified by the Cugnot cart of 1769, unanimously recognized as the first self-propelled vehicle.

Figure 8.1 shows a drawing of the transmission mechanism between pistons and driving wheel. The sole driving wheel is positioned in front of the vehicle and can be steered along with the steam cylinders and boiler (not represented); no suspension is applied.

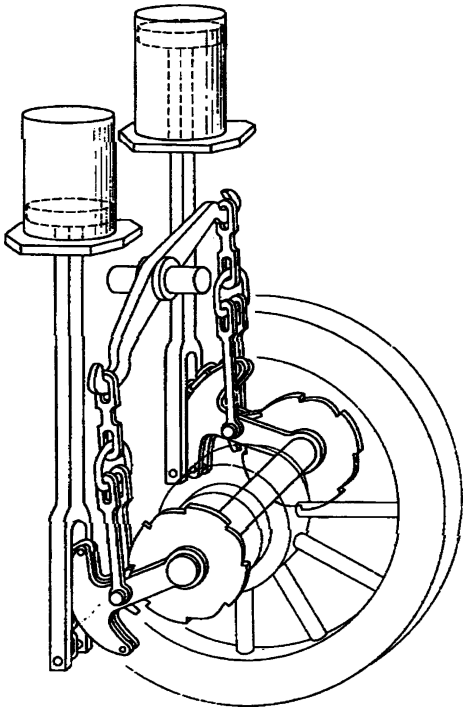


FIGURE 8.1. Drawing of the transmission mechanism from the piston to the wheel in the steam engine cart designed by Cugnot in 1769.

Engine rods are directly connected to the wheel through ratchet gears that convert reciprocating motion of the piston into wheel rotation.

This direct link, eventually made, more simply, with a crank mechanism, was also applied to other steam coaches and used widely on steam train locomotives.

The first operational internal combustion engine vehicle was probably introduced by De Rivaz in 1807. It adopted a similar transmission and exploited the ratchet gear mechanism to keep the vehicle moving under the action of inertia, during the non-useful strokes of the piston.

8.1 MANUAL GEARBOX

Patent documents of 1784 give evidence that Watt foresaw the use of a constant mesh gear box, with dog clutches, to improve the traction performance of a steam engine; it is rather difficult to demonstrate how this idea might have influenced the design of ensuing cars, which did not follow this scheme indiscriminately.

The first commercialized internal combustion engine cars are, without doubt, those born of the efforts of Benz and Daimler in 1885 and 1886; the transmission problem was solved using a scheme completely different from that proposed by Watt.

The complete Daimler car transmission is shown in Fig. 8.2. There are two different transmission ratios made with leather belt transmissions with pulleys of different diameter; the belts are always wound on their pulleys, but the motion is transmitted by only one of the couples, when one of the two tensioners 55 (for the first speed) and 56 (for the second speed) sets its belt at work.

Belt slip capacity is exploited, when the tensioner is not completely engaged, to start-up the vehicle when stopped.

Wheel suspensions are missing and the driveline is consequently simplified.

Many improvements of this scheme were applied to the following Benz car. The two speed transmission was inspired by contemporary workshops, where a single steam or water engine moved a number of working machines. This kind of transmission was probably invented by Anderson in 1849.

The two driven pulleys (in the center of the lower drawing in Fig. 8.3) are coupled with as many idling pulleys (at the outside faces of the driven pulleys); these have a slightly smaller diameter than the driven pulleys and the active cylindrical surface of the driven pulleys is rounded to join the active surface of those idling.

The two driving pulleys (in the back of the car, aligned with the engine crankshaft) have an adequate width to bear the belt on both driven and idling pulleys. Two tensioners can shift the leather belts from matching driven pulleys to matching idling pulleys.

The belts are crossed in order to increase the winding angle on the pulleys; belt tension is adjusted periodically, by changing the distance of the center lines.

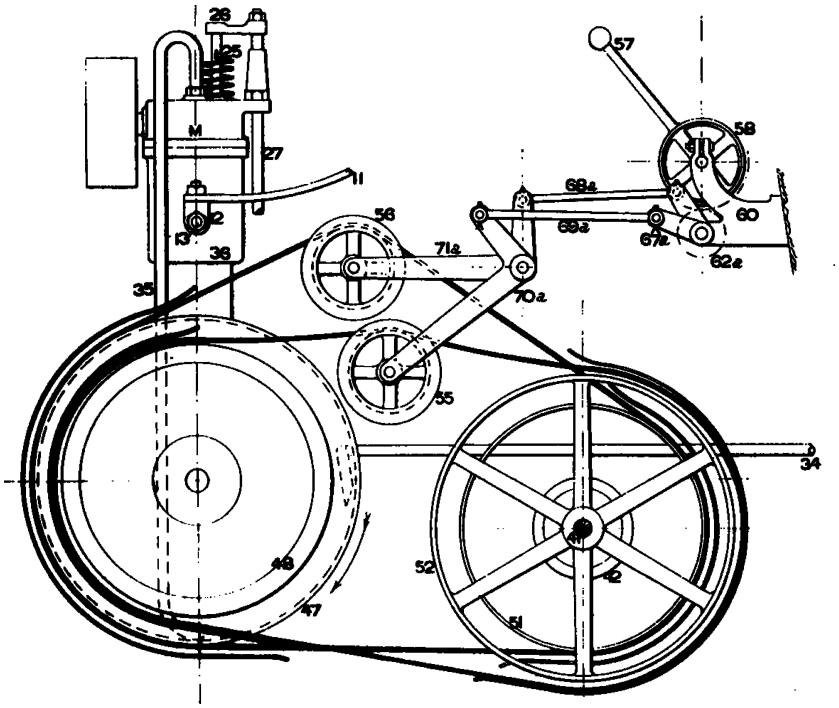


FIGURE 8.2. Complete transmission of the Daimler car of 1885. Two transmission ratios are available, made with a leather belt transmission. The slipping capacity of the belts is exploited to start-up the car.

The rounding on the surface of the driven pulleys makes belt shifting easier; the lower diameter of the idling pulleys decreases the working tension of the belts when they are not working, retarding the need for further tension adjustments.

The start-up function is again performed by exploiting the belt slip capacity; the almost vertical motion of the driving wheels, already using a suspension, is compensated for by two chain transmissions that connect the driven pulleys with the rear wheel hubs.

It should be noticed that the literature on these and other cars of this time did not suggest a sequential use of gears to accelerate the car, but starting-up was allowed with both gears, the choice bound rather to the desired cruise speed, than to the necessary traction force.

The idea introduced by Watt would be applied on next generation design. This is usually noted by saying that the first cars demonstrated how the technical skills of their inventors were polarized on engines, sometimes neglecting the subject of other achievements that had been already obtained in the contemporary state of the art. This fact could be also explained by the difficulty of exchanging ideas through the technical communities of diverse countries.

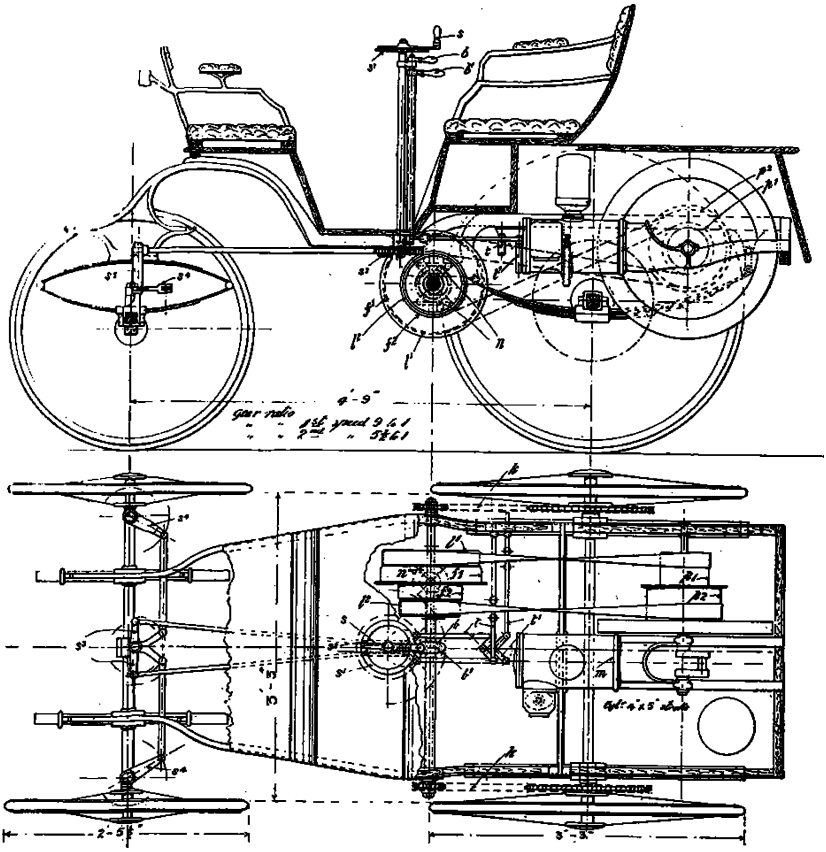


FIGURE 8.3. Two speed transmission of the Benz car in 1886, leather belt type. Two tensioners can shift belts from a position engaged with a driving pulley to that engaged with an idling pulley.

A typical example, similar to contemporary situations, is represented in Fig. 8.4, where a drawing of the FIAT 8/16 HP of 1902 is shown.

The gearbox features three speeds and a reverse; in the oldest cars the reverse speed was missing.

The engine rotates the lower shaft, at the left, a view that represents a side view of the gearbox. This gearbox has a single reduction stage, where all driving tooth wheels are aligned with the input shaft and all driven wheels with the output shaft; on the engine side, we first see the reverse gear, followed by first, second and third gears.

An idle tooth wheel can be noticed, used to reverse the rotation speed of the output shaft on the cross section at the left of the side view.

Driving wheels always mesh with driven wheels and must, therefore, be idle in relation to the output shaft; we will see later how they are engaged. The output shaft rotates a bevel gear, which moves two shafts through a differential.

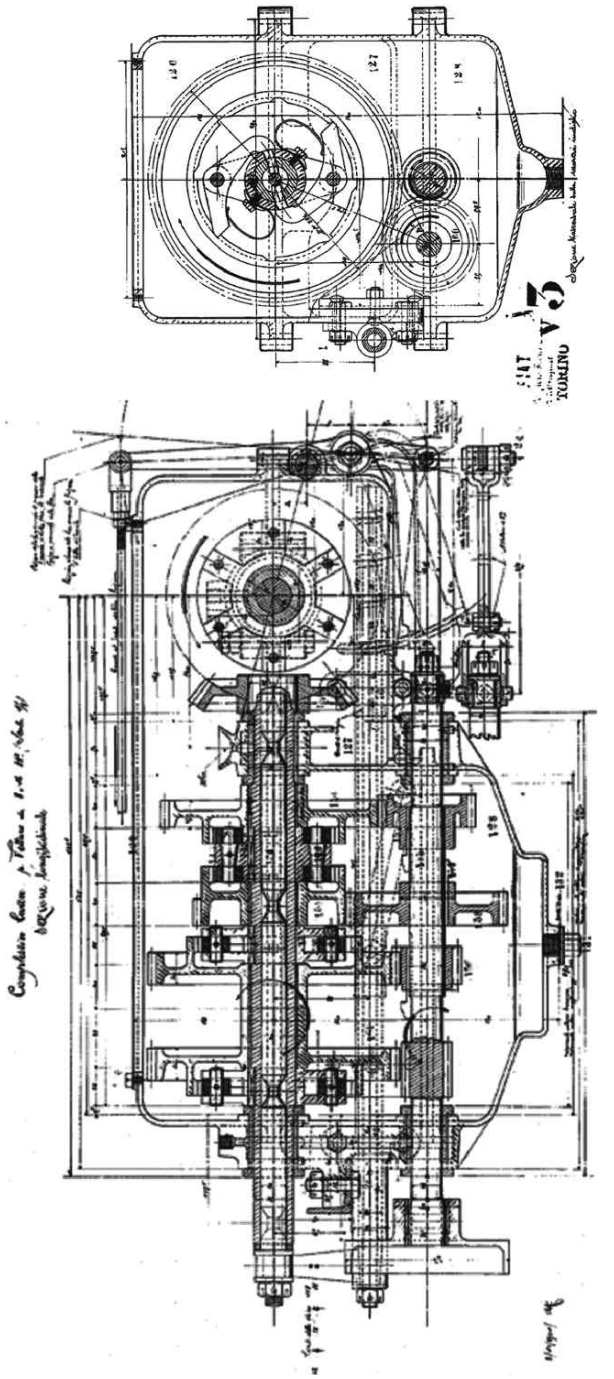


FIGURE 8.4. Single stage, three gear gearbox of the FIAT 8/16 HP of 1902. The internal shifting mechanism is made by a cylindrical rod, mounted inside the driven shaft.

These are coupled through drive chains to the rear driving wheels, according to the scheme already commented on in the previous figure.

A tapered friction clutch, not represented in this figure, allows vehicle start-up and separates the gearbox from the driveline during gear shifts (see for reference Fig. 8.12).

We can notice three new mechanisms not yet introduced: Reverse gear, differential and friction clutch. Their origin is surely older than this car. The reverse gear was introduced by Selden in 1879, the differential by Pecquer in 1827 and the tapered friction clutch by Marcus in 1885.

During these years these three mechanisms were integrated in a transmission suitable for automotive application.

Let us now look at the shifting mechanism. It is rather sophisticated, made by a shiftable cylindrical rod mounted inside a cavity in the driven shaft. This rod shows some annular narrowing, which in certain positions allows two ratchets to engage with the driven wheels; the detail of the reverse gears ratchets can be seen in the cross section on the left.

When one of the grooves faces a pair of ratchets, two leaf springs provide for their engagement with the ratchet gear; when the rod is shifted, ratchets retract, leaving the wheel again idle. The position of the grooves is congruent with a sequential shift stick, where the positions of reverse, first, second and third follow each other.

This kind of gearbox architecture (constant mesh gears) is applied to many cars of that era, but will be soon abandoned because of its complexity and consequent fragility. The next scheme is that of *sliding gear trains* (*train balladeurs*, in french). This invention is older than the automotive world and born outside of it; its emergence was the work of Griffith in 1821.

It is not the first time that a better performing architecture (constant mesh) is abandoned in favour of a less evolved alternative (sliding trains), because a particular component (dog clutch, synchronizer) has not been developed yet. Sliding trains will soon be abandoned again, in favor of constant mesh, when the technological improvements make it possible.

Sliding trains can be exemplified by the FIAT 60 HP car of 1904, whose gearbox is shown in Fig. 8.5.

The gearbox is still of the single stage type and presents four speeds and a reverse speed; tooth wheels, grouped in two trains, from the first to the fourth, can slide on the upper driving shaft. The engine (not shown) is on the left, while on the right we can see bevel gears moving the pinion of the chain transmission through the differential.

The two trains integrate, respectively, first and second speed wheels, and third with fourth; these trains are mounted on a square section shaft that allows them to turn with the shaft, being free to shift along it.

Train sliding is accomplished by suitable sleeves that bring one wheel at a time to engage with its counterpart; the gears mesh only when their correspondent gear is selected.

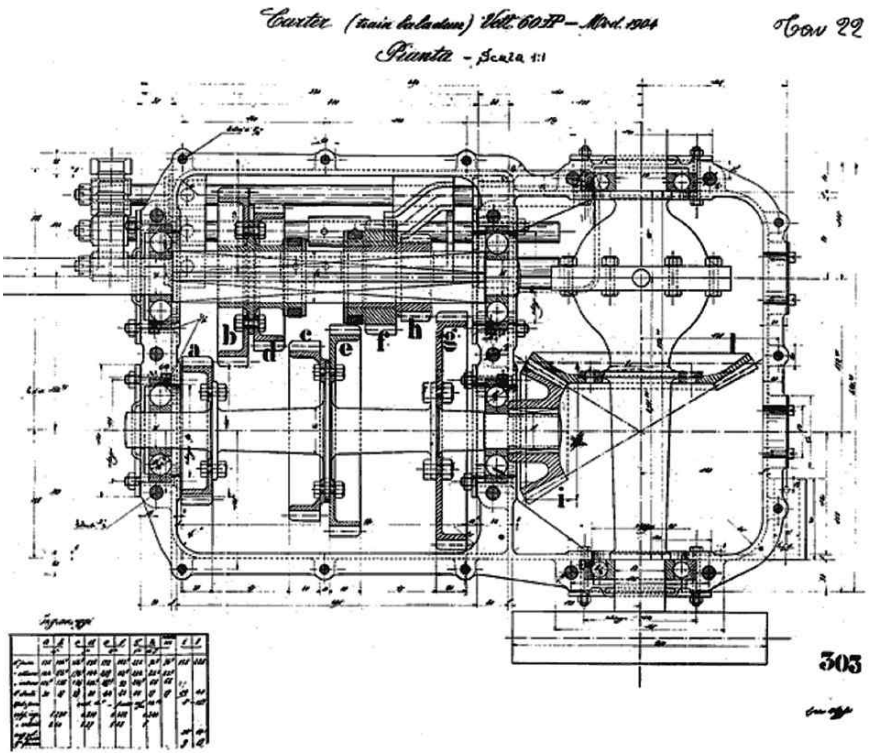


FIGURE 8.5. Sliding train gearbox with four speeds of the FIAT 60 HP of 1904. The two gear trains are made by the driving wheels of the first and second speed and by those of third and fourth speed.

Reverse speed is accomplished by a train of idler gears only, not represented in this figure, meshing with the first speed gears, when their train is in the idle position.

The sleeves are moved by forks fixed to sliding rods, partially visible under the gearbox shafts; there are two rods for moving the first and second train and the third and fourth train, while an additional rod is dedicated to reverse speed.

This kind of lay-out has created the need for selection and shift control sticks, where gear shifting is no longer sequential, but is characterized by two separated motions: one across the car, to select the train to be engaged, the second along the car, to engage the specific gear. This layout will consolidate in practice, surviving unchanged in our contemporary cars.

Sliding train gearboxes required a particular skill of the driver, who had to synchronize the wheels by means of the engine during idling, before engaging the next gear (double clutching or *débrayage double*, in French); an imperfect maneuver was indicated by scratching of the gears in contact at an inappropriate speed.

Some manufacturers improved this architecture, actually returning to Watt's idea of constant mesh gears; an industrial vehicle gearbox of the end of the 1910s is represented, to show this new lay-out, in Fig. 8.6.

In this gearbox three speeds are available and a reverse speed; the gearbox is connected to the engine on the right of the figure, while the output flange is on the left. The flywheel on the output shaft is not part of the engine, but is the drum of the transmission brake.

Input and output shafts are now coaxial; this lay-out is appropriate to the application of a universal joint transmission, beginning to be applied in these years.

Tooth wheels are in constant mesh and have at their flanks a dog clutch; sliding trains are now reduced to the pairs of dog clutches.

The scratch problem is not solved, but possible damages are localized on an auxiliary component, the clutch, that can be sacrificed, with smaller impact on the operation of the gearbox. Dog clutch teeth can also be rounded, making shifting easier, without penalty for wheel dimension.

We can observe that input and output shafts are coaxial, the upper shaft made by two parts, free to have different rotation angles. The left sliding train can alternatively engage the two parts directly or move a third shaft (*countershaft*), below in the same figure.

Only with this dog clutch engaged is it possible to obtain first and the second speeds, with the second train in the figure; with the last train it is possible to engage the reverse speed, connecting the countershaft to the output shaft and then to a pair of idlers (in practice a second countershaft), partially visible on the right of the figure.

Sliding train motions are operated by front cams, that allow in this case sequential control.

A similar architecture, but for control, is still present in front engine, rear wheel driven vehicles.

We should not fall into the mistake of thinking that at this time these technical solutions were consolidated; as often occurs at the dawn of a new technology, deviations from what, for us, is the primary evolutionary path were many. We can see that, after the engine, the gearbox was the preferred subject of innovation of the first automotive engineers.

A comprehensive classification of all attempted solutions is well outside the scope of this chapter, even limiting the discussion to manual transmissions; we will only consider what are in our view some of the more original solutions.

Between 1924 and 1938 Frazer Nash in England built different cars, all with sport performance but affordable price in their category. Essential to these cars was the driveline, as can be seen in Fig. 8.7.

The gearbox is made from chain transmissions, three for forward speeds, one for reverse speed. Driving chain wheels (at right in the drawing) are moved by a bevel gear box, connected to the engine; this bevel gear box includes no differential gear.

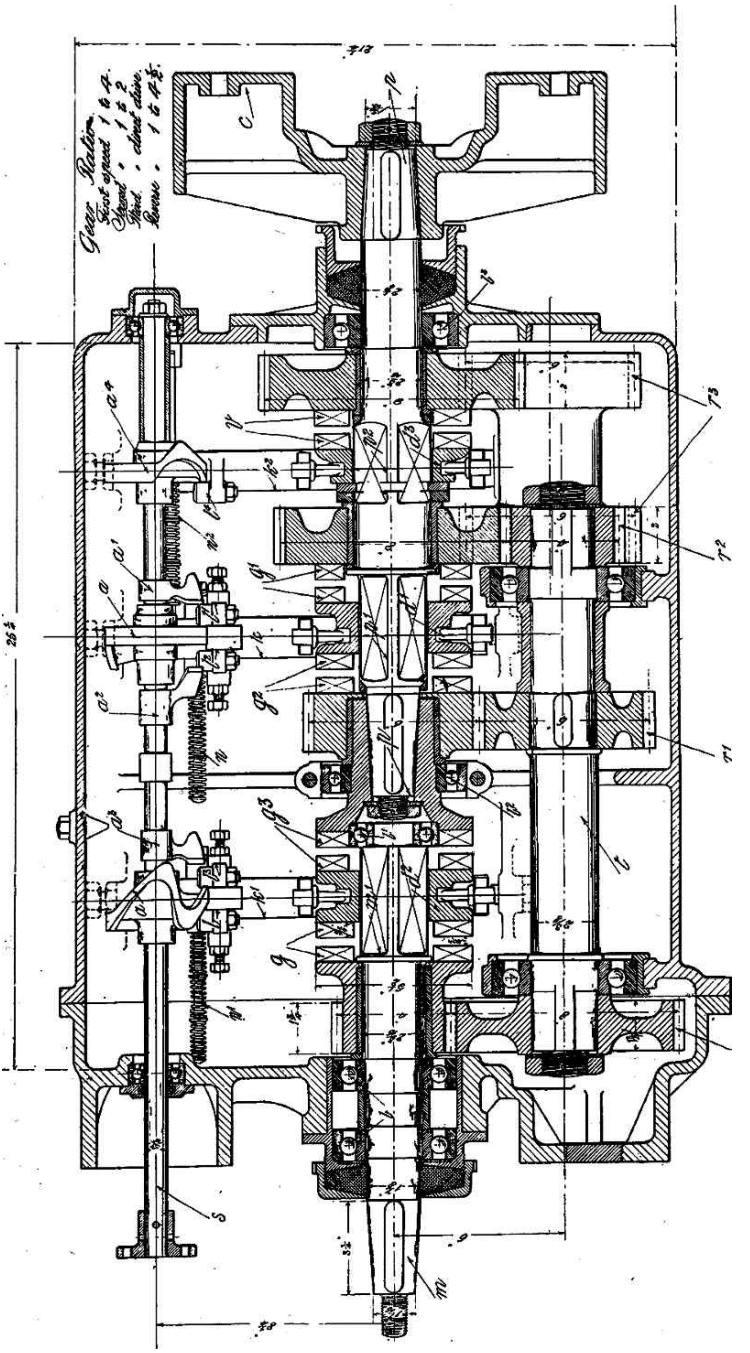


FIGURE 8.6. Dog clutch truck gearbox of a vehicle made in the 1910s. Sliding trains are now reduced to small front tooth gears working as clutches.

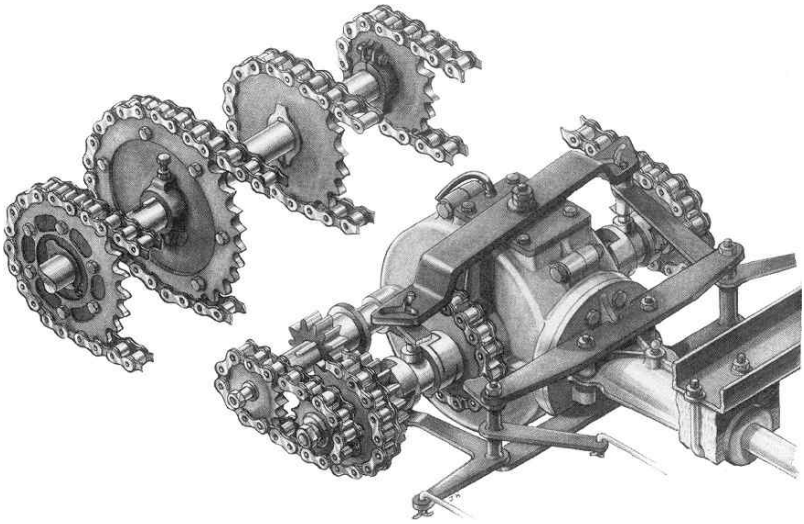


FIGURE 8.7. Frazer Nash gearbox of 1924, made by chain transmissions only; three speeds and a reverse are available. The driven shaft can be adjusted to obtain good alignments and to compensate for chain wear.

On the right half shaft of this intermediate shaft are fixed the sprockets for first speed and reverse; on the left half shaft are sprockets for second and third speed.

Notice the idler reverse system with chain and gears. Driven wheels are directly mounted on the rear one piece shaft; driven wheels can be moved along their shaft for chain alignment.

The lack of the differential made the car difficult to drive, but very manageable for an expert driver on unpaved roads, common in those times.

A chain transmission was also used for compensating for suspension motion.

Engagements were made by simple but sensitive dog clutches. Admirers of these cars extolled the easy road holding, the excellent gearbox maneuverability and the ease with which broken parts could be replaced in the transmission.

The entire mechanism was lubricated by grease in the open air. We can also notice the rocker arm, on the bevel gear box, with particular slots, that match the engagement forks; this mechanism avoided the simultaneous engagement of two speeds, significant considering the effect of centrifugal forces and vibrations.

A car with different technical details, also original and uncommon, is the Sizair Naudin of 1907, also known for featuring one of the first independent suspensions.

Here the design goal should have been to contain gearbox cost, or at least the number of tooth wheels, at that time quite expensive.

The gearbox architecture in Fig. 8.8 can be assimilated to a sliding train gearbox. Driven wheels are reduced to one, with front teeth R. The only sliding train, made by wheels P1, P2 and P3, is installed on a bearing swinging around

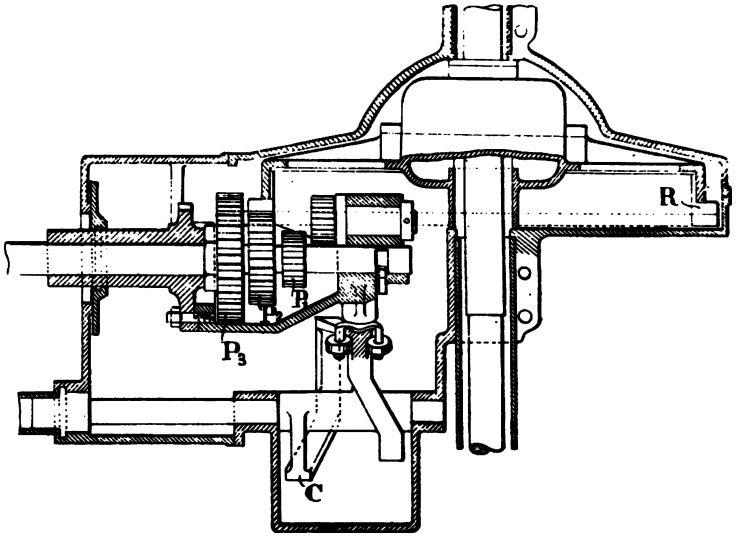


FIGURE 8.8. Sizair Naudin gearbox of 1907. This design can be assimilated to a sliding train gearbox, with bevel gears. The single sliding train made by wheels P1, P2 and P3 is mounted on a swinging bearing to mesh the driving shaft at different distances with the big output front teeth wheel.

the lower shaft in the drawing; the swinging motion of the bearing is necessary to mesh the driving wheel in use with the front teeth wheel, at different distances, depending on the driving wheel diameter.

The swinging motion of the driving shaft is compensated for by a universal joint transmission between engine and gearbox. The cam C at the bottom of the drawing combines shift and swing motions and bears gearing forces.

The spur wheel should instead be bevel gears, for correct matching; the teeth are, nevertheless, approximated with cylindrical teeth, accepting contact errors.

The reverse idler is also present on a dedicated shifting train.

A last example of amazing engineering ingenuity is given by the Turicum of 1904 (probably the only automotive trademark in Switzerland) shown in Fig. 8.9, with a picture of the complete chassis: Here we see one of the first continuously variable transmissions in the history of the automobile.

Nor does this care have a differential; the motion of the rear axle is transmitted with two friction wheels C and D, the first made of solid iron, the second with a rubber tread on its rim. The wheel D is fixed on the shaft E with a spline and groove, that allows the wheel to be shifted along the shaft; the shaft E rests on a swinging bearing G and maintains wheels D and C under pressure with the spring J.

By pulling the lever q it is possible to change the contact point between the two pulleys, between the center of the wheel C (infinite transmission ratio: transmission idling) and its rim (transmission ratio about 1:1). In the idle position,

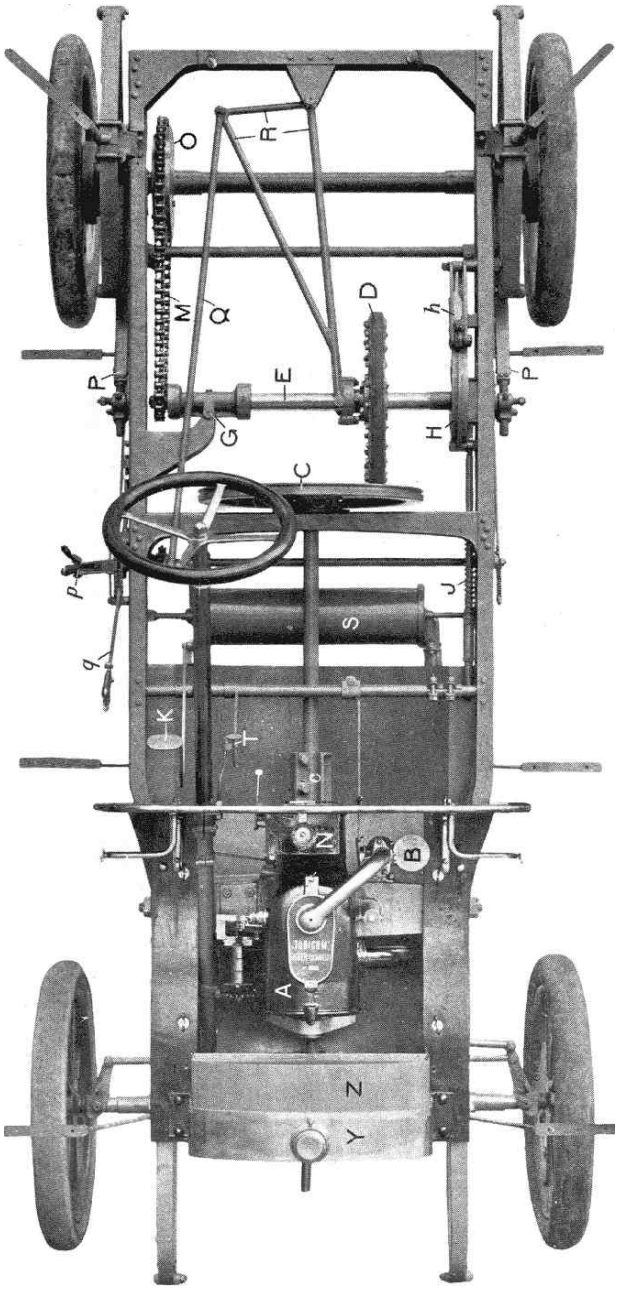


FIGURE 8.9. Turicum chassis of 1904; here we see a continuously variable transmission based upon two friction discs C and D, where the first is made of solid iron and the second has a rubber thread to improve contact friction.

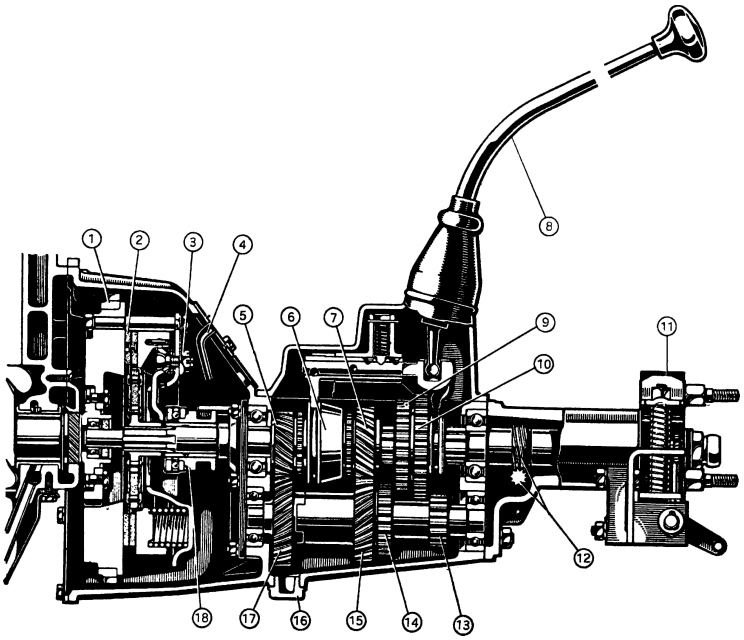


FIGURE 8.10. Four speed gearbox of the FIAT Balilla of 1934. The drawing shows a longitudinal cross section; the rear of the gearbox has a sliding train for first, second and reverse speeds; the front features synchro-mesh gears for the third and fourth speeds.

the friction is eliminated by unloading the spring J; the possible slip between the two wheels is used to start the vehicle up.

These proposals were not imitated by other manufacturers and were probably abandoned by their own inventors. The evolution of manual gearboxes concentrated on perfecting a countershaft or double stage architecture, that became universal on all cars with front engine and rear drive.

An example of this evolution is offered by the gearbox of the FIAT Balilla of 1934 (four speed version, the first gearbox of this car in 1933 was a three speed design) shown in Fig. 8.10.

This gearbox shows two different sections: The rear, for first, second and reverse speeds, features a sliding train with cylindrical straight teeth. The front features helical gears (always meshing) and synchronizers.

This compromise is justified by the high cost of synchronizers, considered high technology components at the time. Synchronizers are limited to the more frequently used speeds, which can also benefit of helical gears, with gearing noise reduction.

On this subject, we notice that engineering manuals of this time suggested, as a good engineering practice, to design the top ratio (in this case the final differential ratio) with values slightly higher then that ideally necessary; this rule was addressed to limiting the number of gearshifts necessary to maintain

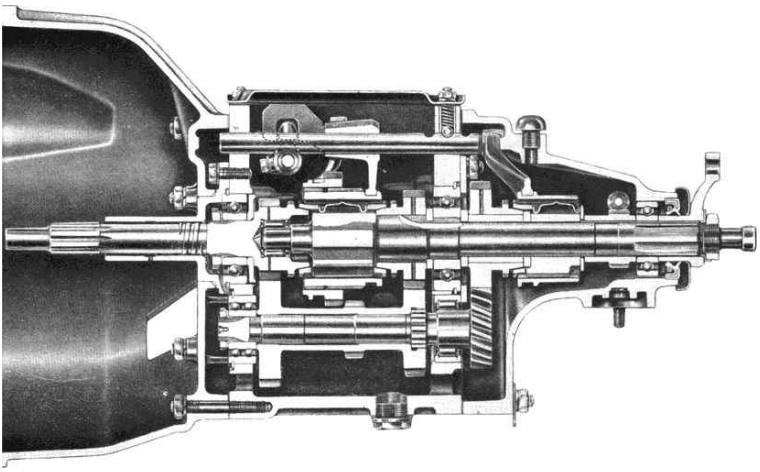


FIGURE 8.11. Four speed gearbox of FIAT 1400 of 1950, with full synchronization, except for the first gear. The reverse speed is made by a sliding idler, not shown in this figure.

the car at cruise speed, and demonstrated the difficulty for drivers to change speed with sliding train gearboxes.

We can therefore assume that synchronizers brought benefits not only for shifting quality, but also for noise and fuel economy.

The gearbox of the FIAT 1400 of 1950, shown in Fig. 8.11, adopts, as do many cars of this time, synchronizers on all speeds but the first, again for reasons of economy; the first is included in a sliding train mounted on the sleeve of the third and fourth speeds. The reverse speed is made with an idler, not shown in the picture, meshing with the first wheels when they are in neutral position.

During the ensuing period, synchronizers were improved and made less expensive thanks to higher volumes; since the 1970s also economy cars have received synchronizers on all forward speeds.

8.2 FRICTION CLUTCHES

An important transmission component, the friction clutch, or, simply, *clutch* posed many problems to designers about operating force and endurance.

In the earliest cars, belt transmissions integrated the clutch function into the gearbox. As we have already seen, bevel clutches were known even at the beginning of the automobile era.

We see on Fig. 8.12 an example of bevel clutch of the first years of the past century.

The friction surface is covered by a leather lining, riveted on a bevel pulley of cast iron; although Frood had already invented the famous synthetic material called Ferodo in 1897, it became widely applied only in the 1920s.

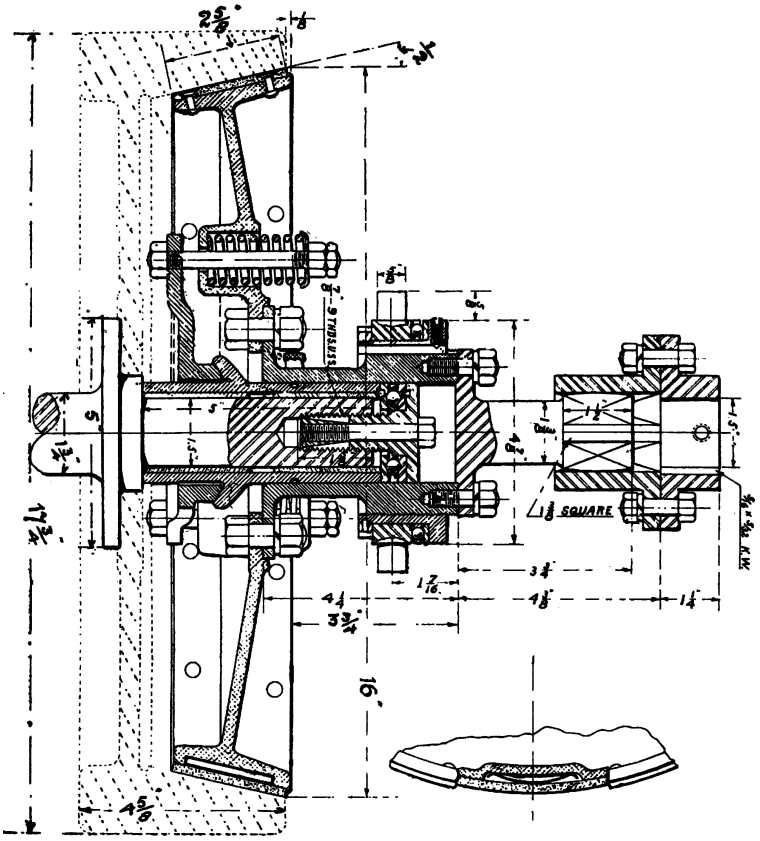


FIGURE 8.12. Bevel clutch with leather linings. Clutch conicity (1:2; lower values could reduce pedal reaction) is limited by the available friction coefficient to avoid irreversible sticking of working surfaces after engagement.

Leather has a friction coefficient similar to modern friction materials, but offers limited performance in terms of heat wasting capacity and endurance; these facts implied a large sizing of active surfaces. On the other hand, leather was, in those times, cheap and easily repaired or replaced.

The active surface of this application is single and shaped like a cone; this shape was chosen to limit the disengagement force on the pedal, which depended upon the cone diameter and engine torque. Due the difficulty of integrating the leather lining into its support disc, the active surface was usually single rather than double, as it would be in modern clutches.

If we refer again to the previous figure, we can notice that the engine flywheel features a short shaft bearing the reaction structure of the load springs working on friction surfaces.

Many coil springs (only one is shown in the cross section) push the tapered friction disc into the flywheel. The leather lining is riveted on this disc; very thin

leaf springs are set between the lining and the disc to make the engagement more progressive.

With this kind of architecture, much different from contemporary methods, the gearbox input shaft must be able to slide on a square counterpart.

Friction conicity (1:2 on this drawing) is limited by the friction coefficient between leather and iron, in order to prevent irreversible sticking of the clutch after engagement.

The large engine displacement of many cars and the limited dimension of the flywheel made many clutches too heavy to be operated; for this reason other mechanisms were also developed.

The idea was to exploit the mechanical property of wound linings to reduce working forces. Here the friction force is itself used to increase contact pressure, as through the leading shoes of drum brakes. This principle was applied through band clutches.

An application of this principle is shown in Fig. 8.13; a coil spring with rectangular section is installed in a cavity in the flywheel; the coils are quite close to each other. One end of the spring is fixed directly to the flywheel, through the eye in the lower part of the figure; the other end is instead connected by a rocker arm D.

If an ogival body is moved closer to the rocker arm, it is possible to twist the spring and reduce its internal diameter.

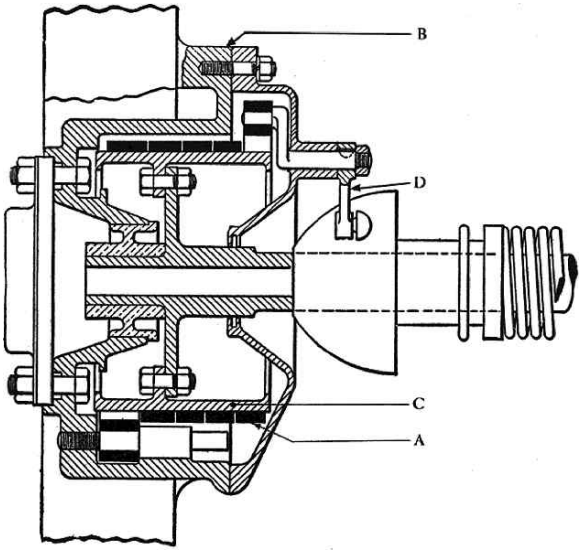


FIGURE 8.13. Clutch with spiral friction spring. The clutch pedal moves the ogive that closes the spring onto the input shaft, creating a friction force; the same force increases the band tension. The resulting friction torque is a function of the initial tension through an exponential function of the winding angle of the spring.

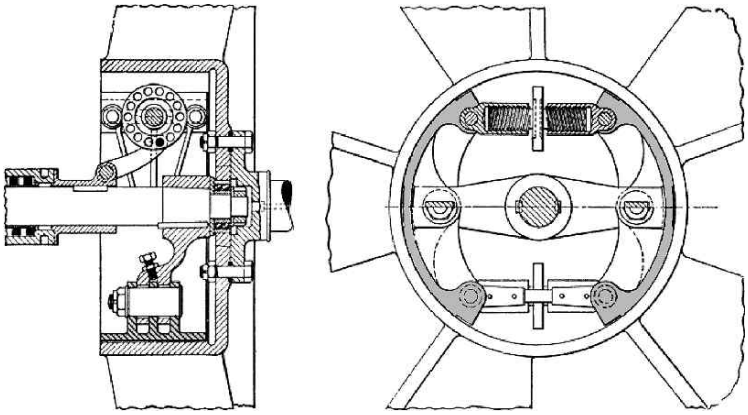


FIGURE 8.14. Radial shoes clutch. The shoe displacement is caused by a screw mechanism, operated by the clutch pedal through a crank.

The gearbox input shaft C is surrounded by the spring with a slight play; when the ogive is advanced through the clutch pedal the spring closes on the shaft with a resulting friction torque.

The friction tension along the spring coil increases the tangential tension toward the eye, without increasing its reaction on the rocker arm; the resulting friction torque is an exponential function of the winding angle, which can be increased indefinitely.

With a modest friction coefficient between metals it is possible to transmit the desired torque with a reasonable force on the pedal; the downside is the difficulty of the engagement maneuver, only partly eased by spring elasticity.

A very different configuration of the same principle is given in Fig. 8.14; the torque is transmitted by two shoes that expand in a drum, as in a drum brake.

The shoe motion is created by a screw that is moved through a crank and rod mechanism; the disc shape of the crank is chosen to allow a simple play adjustment, to compensate for lining wear.

Faced with a difficult problem, inventors investigated many different solutions before consolidating and improving the best one; to solve these problems electric and hydrostatic transmissions were also investigated and applied.

The final solution was consolidated in the 1930s with the single disc clutch with synthetic friction linings; one example from this period is shown in Fig. 8.15.

The friction surface is now flat and double; with the same force it is possible to transmit a double torque. The friction disc is mounted within two surfaces (the flywheel and the pressure disc) that are compressed by a number of coil springs; a set of levers on the pressure disc are used to release the clutch with the axial motion of a thrust bearing.

This kind of clutch received its last improvement by the application of cup springs; these were introduced at the end of the 1970s and allowed many advantages, such as a further reduction of pedal force and a general simplification of the mechanism.

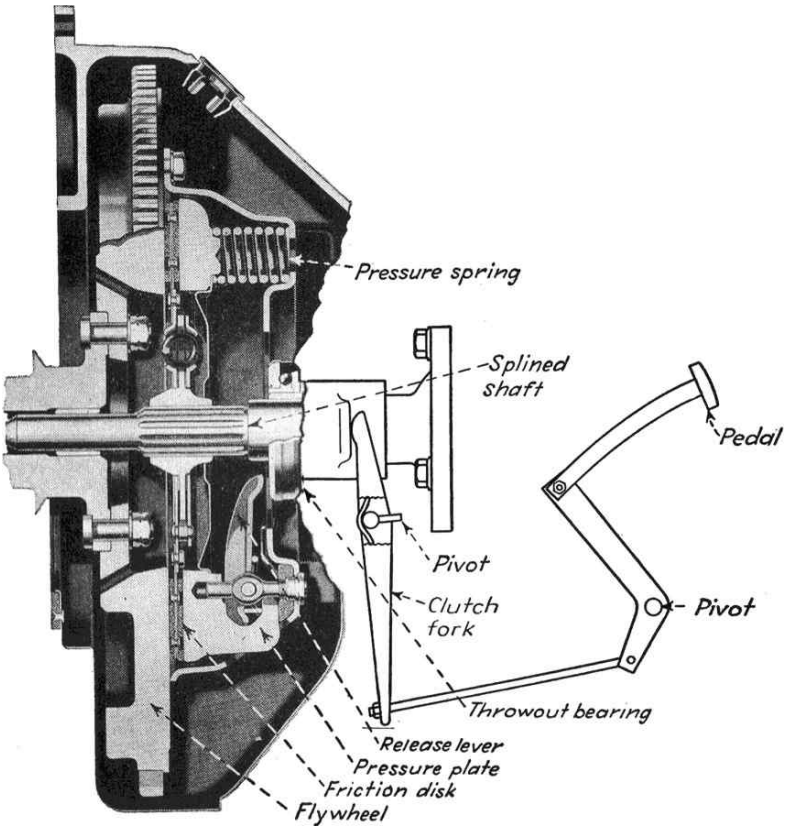


FIGURE 8.15. Dry single disc clutch with coil pressure springs. A set of release levers articulated on the pressure plate is used to disengage the clutch, through the displacement of the throwout bearing.

8.3 AUTOMATIC GEARBOXES

Automatic gearboxes have had their own history, one that received a crucial contribution from the American automotive industry.

We do not suggest that Europe failed to contribute to this development; we will see, in fact, that many fundamental inventions were developed on this continent. Nevertheless the European market, smaller and more fragmented, did not justify the mass production of this gearbox until recently.

The problems to be solved in developing an automatic gearbox included a different mechanism for engaging gears and starting the vehicle, easier to operate with less sophisticated automatic controls. These could be mechanical (exploiting centrifugal forces) or hydraulic (exploiting the pressure variation of the oil in a rotary pump).

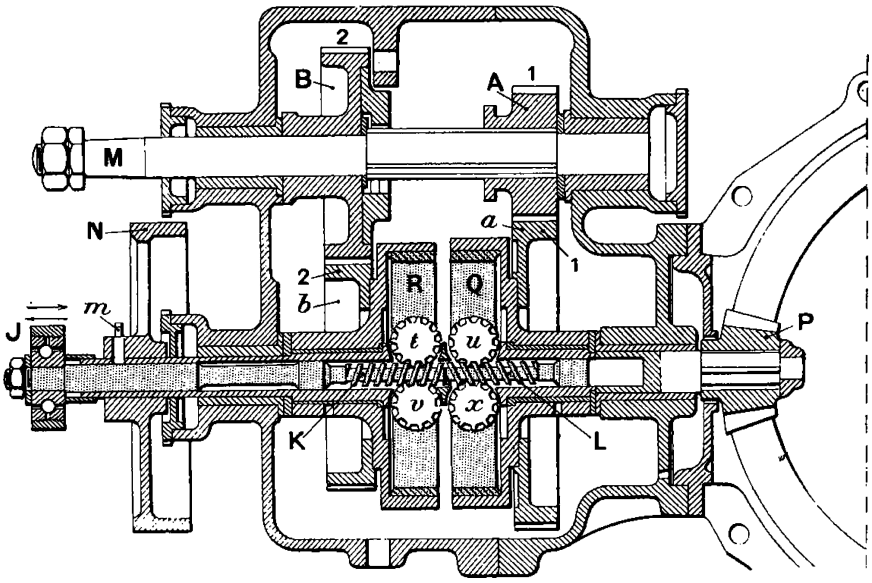


FIGURE 8.16. The De Dion & Bouton gearbox can be considered as a precursor to power-shift gearboxes. By shifting the shaft with the bearing J, it is possible to engage one of the two shoe clutches available on each gear; a start-up clutch is not necessary.

Today this problem appears in a new context, because electronic microprocessors allow easy automatization of synchronizers and friction clutches, in use on manual gearboxes; many existing vehicles already testify to this statement.

The first step was the development of gearboxes where speed shifts were possible without danger to tooth wheels and parts to be synchronized.

From this point of view we can consider as a precursor the manual gearbox of de Dion & Bouton, developed at the beginning of the past century and shown in Fig. 8.16.

This single stage gearbox has but two speeds; we can see at the left upper the input shaft and on the lower right the output shaft, which moves through a bevel gear the pinions of the chain drive.

The two gear always mesh, with the driven wheels idling on the output shaft; the wheel engagement is made by shoe clutches, similar to that already discussed in Fig. 8.14; these are controlled by the screws moving the gears t, v, u and x.

By shifting the shaft with the thrust bearing J, it is possible to engage one clutch and to disengage the other.

In this transmission, a start up clutch is not used during gear shifts.

Although developed for manual gearboxes, this kind of clutch is surely a relevant precursor of the powershift gearbox with band brakes and multi disc clutches.

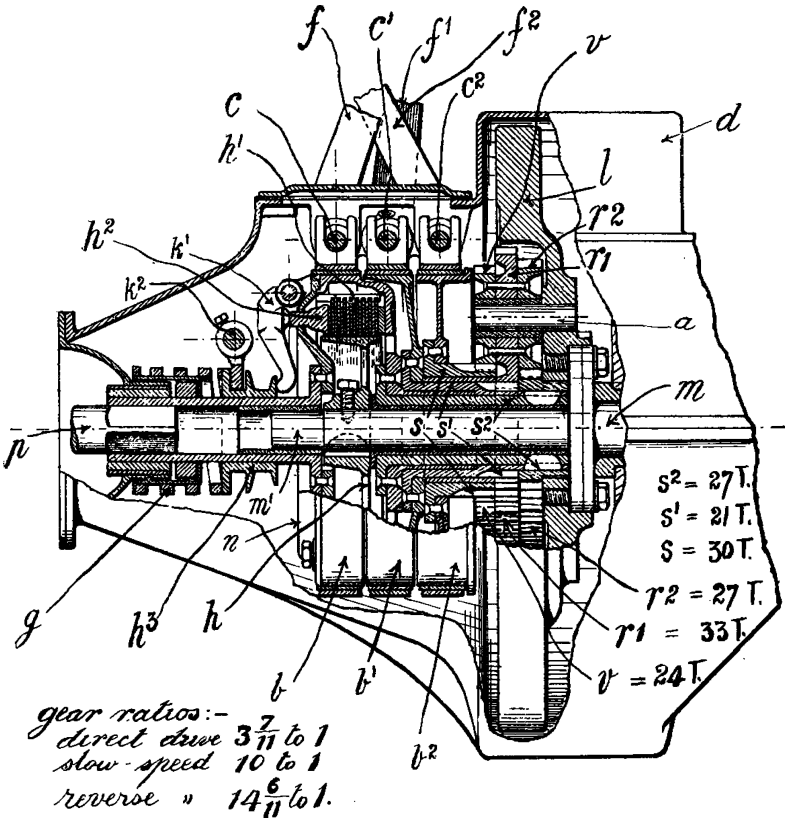


FIGURE 8.17. Epicycloidal gears gearbox of Ford model T of 1908. One reverse, one reduced forward speed and a direct drive are available.

A second gearbox of historical relevance is that of the Ford Model T of 1908, the first car to be produced in the millions.

Figure 8.17 shows a section of this gearbox; it is made with epicycloidal gears, instead of gears with a fixed rotation axis.

These gears did not originate with Ford, for they were already known in other applications. The epicycloidal gearbox may have been invented by Bodmer in 1834, although there is evidence that these mechanisms were already known to the ancient Greeks in applications for astronomical computations.

In this figure we see the three satellites v, r^1 and r^2 (the unusual position for the subscript, not to be confused with an exponent, is drawn from an original figure), rotating on a single carrier, fixed to the engine flywheel. These mesh with the corresponding sun gears s, s^1 and s^2.

If we imagine keeping the sun s stopped, by rotating the flywheel and the carrier we obtain a reduced output speed in the opposite direction, at the sun gear s^2, fixed to the output shaft.

On the other hand, by keeping the sun s^1 stopped, we can obtain a reduced speed in the same direction, again at the sun s^2; in the figure a note reports the transmission ratios that were obtained, including the differential transmission ratio.

If the multi disc clutch h^1 is engaged by shifting the sleeve h^3, it is possible to put the gearbox in direct drive by fixing the hub h, rotating with the crankshaft, with the output shaft.

To obtain the different states of the gearbox, the sun gears rotate with the drums c, c^1 and c^2, which can gradually be stopped with their band brakes; band brake control is performed by front cams, moved by pedals on the car dashboard; the lower part of these pedals is shown in the figure with the letters f, f^1 and f^2.

The pedals have a spring system that makes them stable either in the released or depressed position; each pedal raises if the other is depressed.

When engine and car are stopped, the pedal f must be depressed and the clutch h engaged; in this way the vehicle is in park condition.

By releasing the clutch h, through a lever, the engine is disengaged and can be cranked. The car is still stopped.

By depressing one of the pedals f^1 or f^2, the pedal f is raised, the car is left free to move and will be started-up in low gear forwards or backwards; speed inversions can also be made by a moving vehicle, and start-up on slopes is made easier.

As soon as the suitable speed is reached, by engaging the clutch h the pedal f^1 will be released, obtaining direct drive.

The gearbox is controlled by driver actions, but clutch management is performed automatically during gear shifts.

From this scenario to a fully automatic gearbox the way was long, but these achievements brought the final result closer.

The configuration of this gearbox allows us to understand why epicycloidal gears were preferred to conventional ones for the new automatic transmissions: Because of the ease of integrating brakes and clutches.

A further step was made by Wilson, in England, in 1928, who proposed a gearbox made of two different epicycloidal gear trains in series, in which the carrier of the first gear was connected to the ring of the next gear. With two gears it is possible to obtain three speeds forward, one of them being a direct drive and a reverse speed.

The three speeds were obtained by braking drums with bands, as in the Ford gearbox; a schematic example of the Wilson gear train is shown in Fig. 8.20, in the chapter dedicated to automatic transmissions.

These gearboxes, similar in use to those of the Model T, were semi-automatic with manual preselection; according to this concept, a small lever near the steering wheel was used to select in advance the next gear to be used. At this point no gearshift commenced, but the brake mechanisms were arranged for the gearshift to be made. This occurred as soon as the driver depressed a pedal for this purpose, set in the position normally used for the clutch pedal.

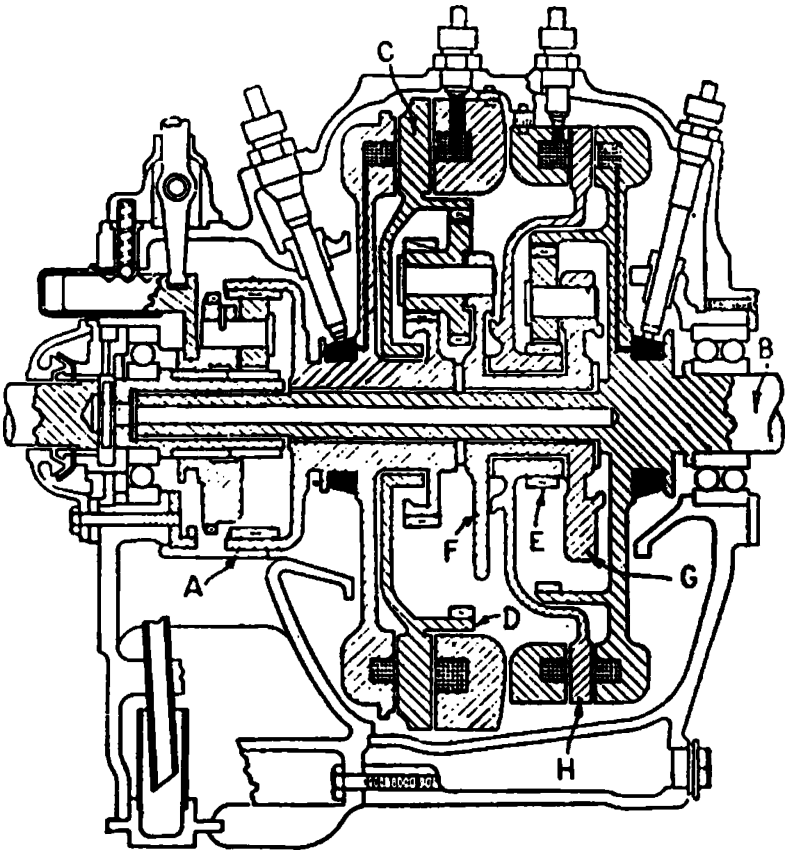


FIGURE 8.18. Semi-automatic gearbox produced by Cotal from 1934. The different elements of epicycloidal gears are braked by electromagnets; the reverse is engaged manually.

The driver was supported, in this way, in executing a coordinated maneuver of the gear stick and clutch; the energy for this function was still produced by driver muscles through a pedal.

A particularly advanced semi-automatic gearbox was introduced by Cotal in 1934, in France; Fig. 8.18 shows a cross section of this gearbox.

This gearbox includes three epicycloidal gears; the engine is on the left, the output shaft on the right.

Toroidal electromagnets can stop elements of the gear train; in particular, the first puts the corresponding gear set into direct drive, by fixing sun and ring gears together; the second obtains a reduced speed. At right, the third electromagnet obtains a faster speed, while the last achieves direct drive.

By energizing electromagnets in combination, two reduced speeds, a direct drive and an overdrive can be obtained.

A small switch with five positions, set near the steering wheel, allowed the four ratios to be obtained automatically, without use of clutches, whose function was controlled by electromagnets timing and inertia of parts accelerated or slowed down during shifts; the fifth position was for the idle gear, with all electromagnet circuits open.

The first epicycloidal gear on the right is operated, instead, manually, when the car is stopped and the transmission in idle position; a control lever moves the carrier back and forth, which can engage with the ring gear, obtaining a forward speed, or can be stopped, obtaining a reverse gear. Vehicle motion can be obtained, after this manual shift, with the first gear, controlled by its electromagnet.

The most relevant inconveniences of this gearbox were its heavy weight and large size.

Semi-automatic Wilson and Cotal gearboxes were used primarily by European manufacturers specializing in luxury cars; the Second World War crisis caused many of these manufacturers to disappear, these transmissions with them.

The final step towards modern automatic gearboxes was taken by exploiting hydraulic torque converters.

The torque converter had been introduced by the German naval industry, after the invention of Föttinger in 1905, well before its application in cars.

He patented a torque transmission system using a centrifugal pump and a turbine, in the same hydraulic circuit. With this device, the torque transmission is obtained by the momentum variation of the flow through the rotating blades, and is also possible when the pump (the engine) is rotating and the turbine (the vehicle) is stopped.

The idea was developed further through the design of an integrated device of reduced dimensions almost interchangeable with the conventional friction clutch.

In 1910, a patent for a hydraulic clutch simplified by the elimination of the reactor element was filed.

Again in Germany, in 1928, the research consortium Trilok developed the homonymous torque converter, able to obtain in a single machine the performance of the torque converter and the hydraulic clutch. This was done by mounting the reactor element with a freewheel.

The first automatic gearbox developed for a car was produced by GM; called the Hydramatic, it has been produced since 1939: A cross section of this gearbox is shown in Fig. 8.19.

In this figure, starting from the left side of the engine, we can see the hydraulic clutch, followed by three epicycloidal gear trains, able to obtain three forward speeds and a reverse speed; engagements and disengagements are obtained by two band braked (37 and 16) and two multi-disc wet clutches (7 and 17).

Brakes and clutches are operated by oil pressure, generated by the gear pump 33 on the front side of the gearbox, and modulated both by servo valves and a manual control on the steering wheel.

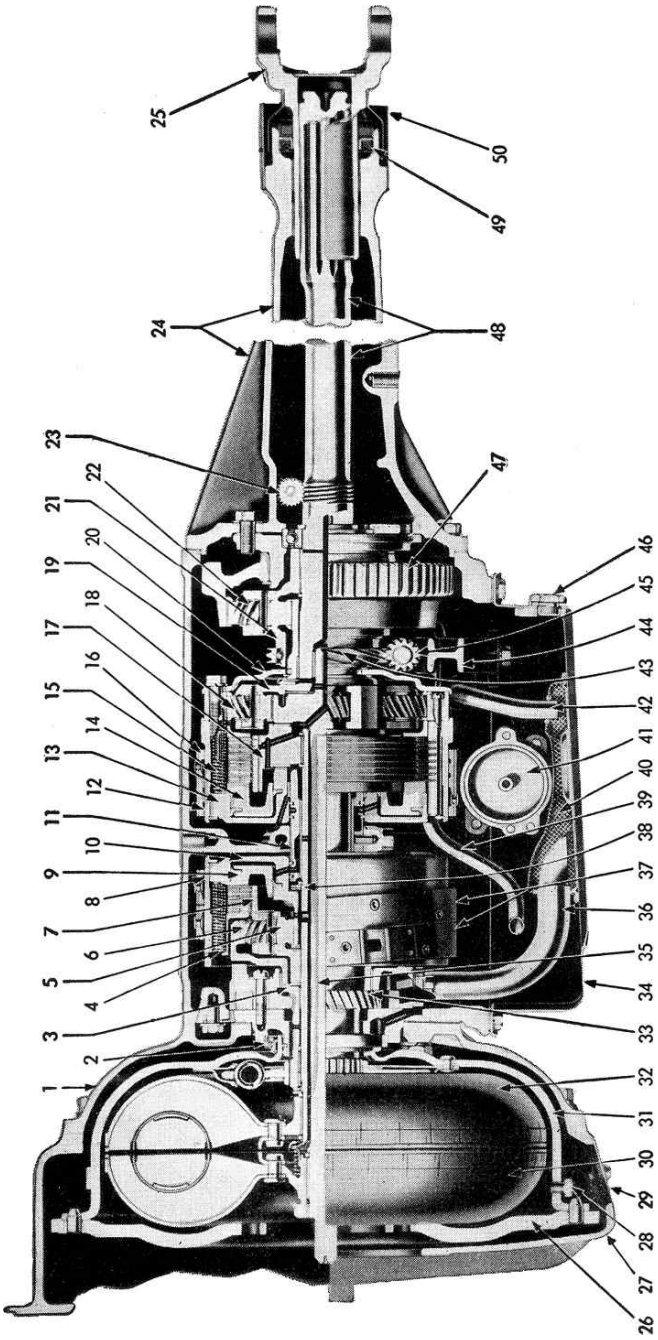


FIGURE 8.19. The first automatic automobile gearbox, produced since 1939, is the Hydramatic by GM. We can see, starting from the left side of the engine, the hydraulic clutch, followed by three epicycloidal gear trains, able to obtain three forward and a reverse speed.

Gear shift automatization is based upon the comparison of oil pressure generated by this first pump (dependent on engine speed) with a pressure generated by a second pump driven by the transmission output shaft (dependent on vehicle speed). The difference between these two pressures is used to move the gear shift servo valve; this valve is also made sensitive to the accelerator pedal position through a spring loaded mechanical link.

This system worked quite well on plain roads, upshifting speeds at higher vehicle speeds with higher accelerator compression; on slopes or on bending roads the automatic control had to be corrected by the manual selector.

We can notice that the hydraulic clutch is always subject to the engine torque; the clutch was used for start-ups and to dampen driveline torque vibrations.

The bulk of these gearboxes was absorbed by war production; only in 1946 did their application on commercial cars commence, to the appreciation of the public.

The Dynaflow gearbox, also from GM, has been produced since 1948. It introduced many improvements over the previous model (Fig. 8.20). The epicycloidal Wilson gear train, much simpler, allowed three forward and a reverse speed to be obtained, with two band brakes and a multi-disc clutch used in combination.

The most relevant step forward was the introduction of a refined torque converter, featuring a two stage reactor on freewheels; with this device it was possible to start-up the car with a torque transmission ratio greater than two (instead of one, by definition the ratio on the hydraulic clutch), allowing in the meantime the torque converter to function as a clutch, with better efficiency when input and output torques on the converter were equal.

This scheme is still present in automatic gearboxes, even if the need for a higher number of transmission ratios has justified the application of additional epicycloidal gear trains.

It is also interesting to remember the automatic gearbox designed in 1949 by the Dodge Division of Chrysler, with very original features.

Figure 8.21 shows the clutch of this gearbox; it includes a hydraulic clutch and a pedal friction clutch in series. Some gear shifts always demand a pedal clutch, but they are rare, thanks to a particular automatization device.

The twin friction and hydraulic clutches allow transmission vibration dampening and a smooth start-up, even if the pedal is released without particular skill; in addition, the car can be kept stopped on a slope simply through the use of an accelerator pedal. The next start-up is considerably easier.

Similar twin clutches were also applied in combination with conventional manual gearboxes on some European cars, such as the Fiat 1900.

The gearbox is explained in Fig. 8.22; it should not be confused with a simple counter shaft gearbox.

The main difference consists in the constant mesh wheel mounted with freewheel that allows torque to be transmitted to the countershaft, but not vice versa, and on a disengagement sleeve of this freewheel on the countershaft; it is

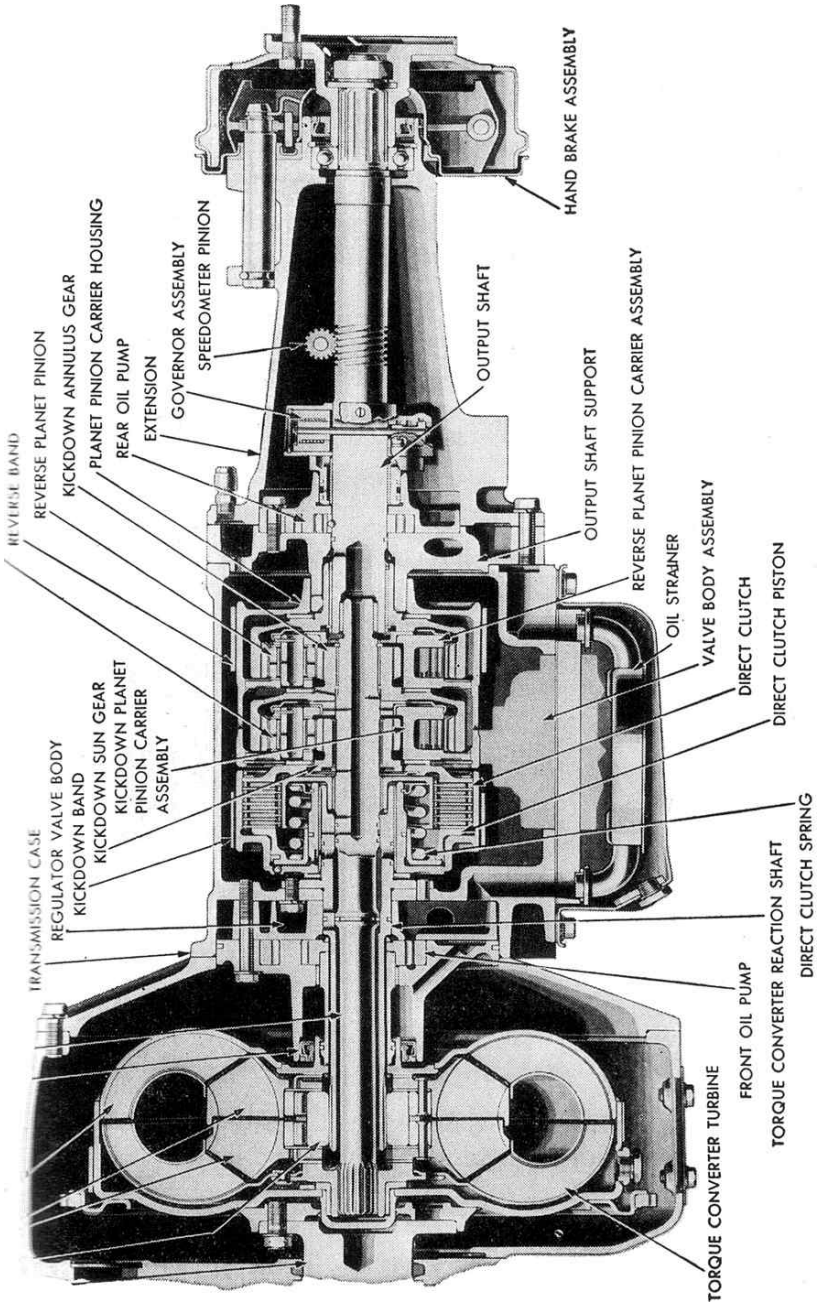


FIGURE 8.20. The Dynaflow gearbox, again by GM, has been produced since 1948 and can be considered an improvement over previous design. The Wilson epicycloidal gear train allows three forward and a reverse speed.

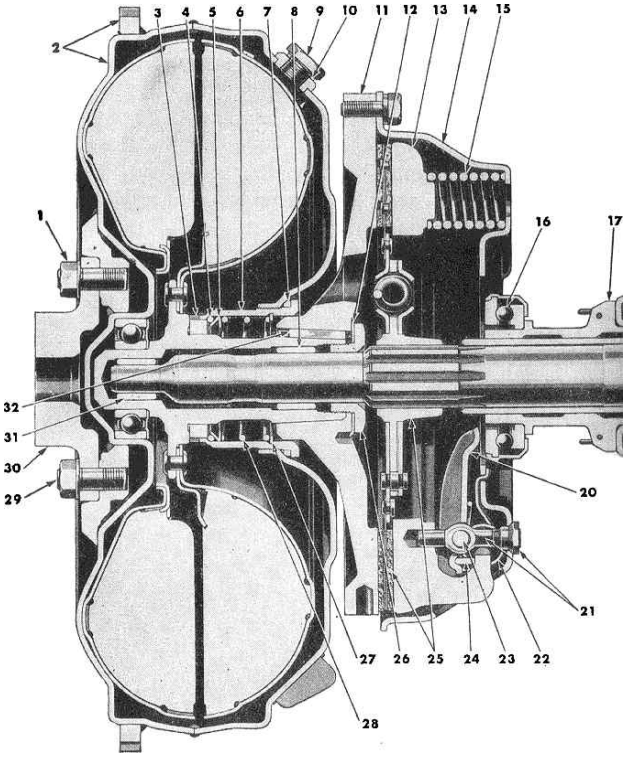


FIGURE 8.21. Hydraulic clutch of the Gyromatic semi-automatic gearbox, produced since 1949 by the Dodge Division of Chrysler; we can also see a conventional friction clutch with pedal control.

in the disengaged position in the first figure on the left. The whole set of figures represents the different gearbox states.

The gearbox control features a manual lever working on the sleeve, on the right of the output shaft, through a suitable leverage; if this sleeve is set to the left, first and second speeds can be obtained; if it is set to the right, third and fourth speeds are achieved.

This maneuver should be made by disengaging the friction clutch. Transmission ratios and engine displacement were such as to justify low speeds on slopes or in urban driving, while the remaining gears were recommended for suburban driving, including related start-ups.

Upshifts from first or third gear and downshifts from fourth or second gear were made automatically, by a tachometer device shifting the freewheel sleeve. Notice that when the gearbox is in the low gears, the third speed gear acts as a separate constant mesh gear.

These speed shifts did not require clutch disengagement because of the properties of the freewheel: The figures represent the positions in first and second

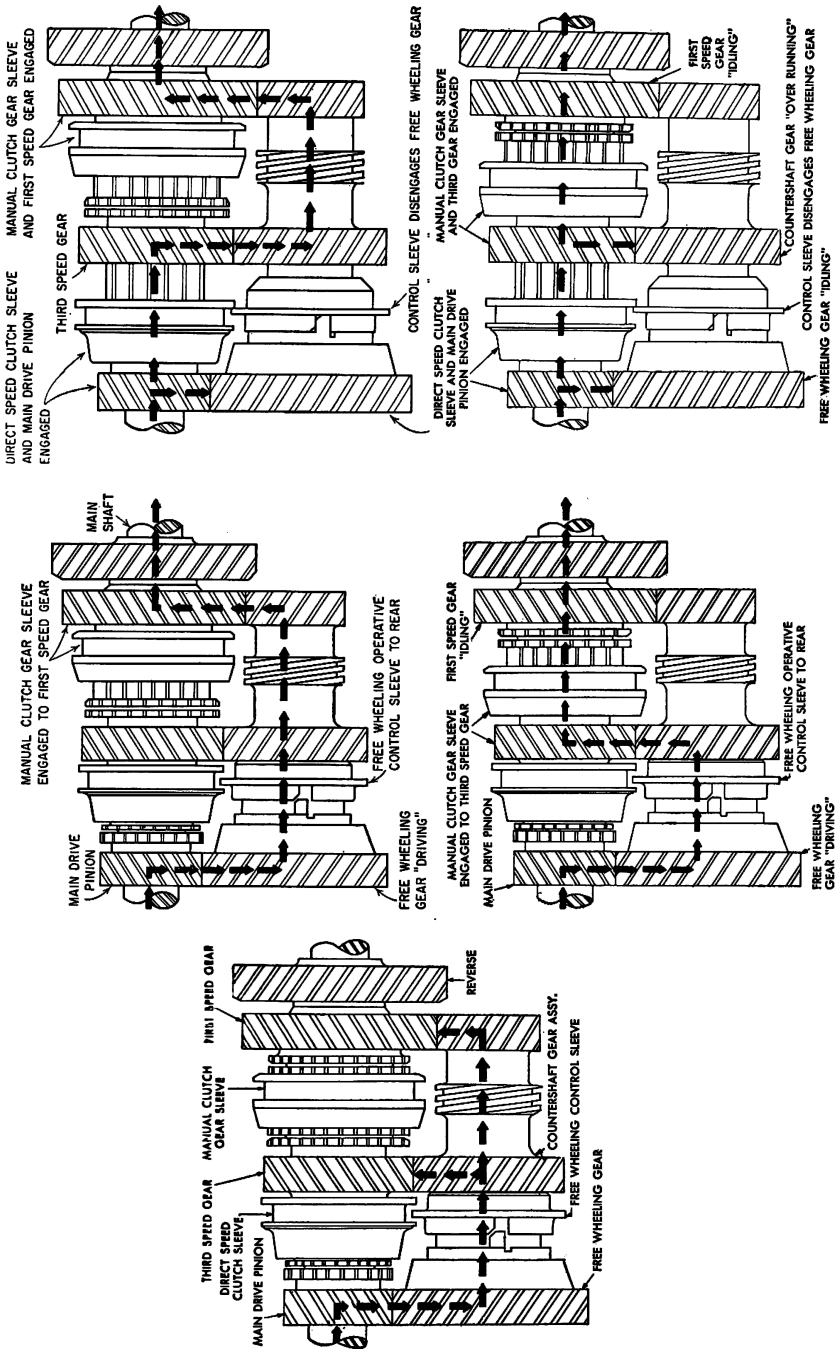


FIGURE 8.22. Scheme of Chrysler's Gyromatic gearbox. The gearbox has a manual control to select a low (first and second speed) and a high (third and fourth speed) speed range. Speed shifts from first to second and from third to fourth and back are fully automatic.

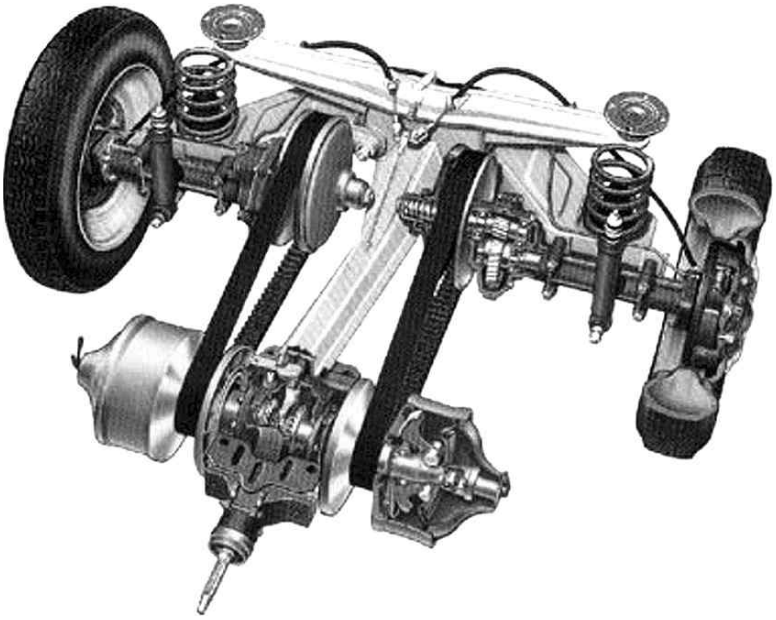


FIGURE 8.23. Automatic Variomatic transmission of the DAF Daffodil of 1950. It is made by expandable pulleys and rubber belt, reinforced by cords.

speed (on the upper row) and of third and fourth speeds (on the lower row). The first figure at the left represents the idle position.

Dotted lines represent the power flow at different speeds; the reverse speed was obtained by shifting an idler to engage the smallest wheel on the countershaft, with the largest on the output shaft.

A particular European contribution to automatic gearbox development was introduced by DAF Daffodil, with the Variomatic transmission of 1950. This transmission was probably the first reliable application of the continuously variable transmission to a car.

This transmission suitable for front engine, rear drive cars is shown in Fig. 8.23; the engine drove two expandable steel pulleys through a transmission shaft and a differential. These pulleys drove similar pulleys connected to the driving wheels.

The sides of the driven pulleys were compressed by coil springs that guaranteed the correct friction with a rubber belt; the sides of the driving pulleys were, instead, compressed by centrifugal masses and engine manifold pressure. Through this device speed ratio variation took into account engine speed and required torque.

A centrifugal friction clutch made car start-up completely automatic.

This transmission received no further application because of its strong impact on vehicle architecture.

The concept was completely reworked by Van Doorne (the DAF holding company), introducing a complete redesign. This study had as its objective the development of a steel belt variable transmission of reduced dimensions, capable of being interchanged with conventional manual gearboxes. An experimental application was made by Fiat and Ford and followed later by mass production.

This kind of automatic gearbox has now received a number of applications on different car brands.

9
MANUAL GEARBOXES

9.1 MANUAL GEARBOX CLASSIFICATION

Gearboxes are normally classified according to the number of toothed wheel couples (*stages*) involved in the transmission of motion at a given speed; in the case of manual vehicle transmissions, the number to be taken into account is that of the forward speeds only, without consideration of the final gear, even if included in the gearbox.

Therefore there are:

- *Single stage* gearboxes
- *Dual stage* or *countershaft* gearboxes
- *Multi stage* gearboxes

Figure 9.1 shows the three configurations for a four speed gearbox.

It is useful to comment on the generally adopted rules of these schemes. Each wheel is represented by a segment whose length is proportional to the pitch diameter of the gear; the segment is ended by horizontal strokes, representing the tooth width. If the segment is interrupted where crossing the shaft, the gear wheel is idle; the opposite occurs if the segment crosses the line of the shaft without interruption. Then the wheel rotates with the shaft. Hubs are represented according to the same rules, while sleeves are represented with a pair of horizontal strokes. Arrows show the input and output shafts.

Single stage gearboxes are primarily applied to front wheel driven vehicles, because in these it is useful that the input and the output shaft are offset; in

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

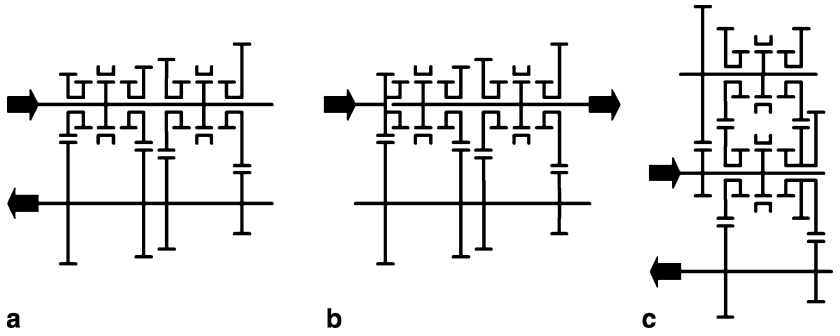


FIGURE 9.1. Schemes for a four speed gearbox shown in three different configurations: a: single stage, b: double stage and c: triple stage.

conventional vehicles, on the other hand, it is better that input and output shafts are aligned.

This is why rear wheel driven vehicles usually adopt a double stage gearbox.

The multi-stage configuration is sometime adopted on front wheel driven vehicles with transversal engine, because the transversal length of the gearbox can be shortened; it is used when the number of speeds or the width of the gears do not allow a single stage transmission to be used.

It should be noted that on a front wheel driven vehicle with transversal engine, having decided on the value of the front track and the size of the tire, the length of the gearbox has a direct impact on the maximum steering angle of the wheel and therefore on the minimum turning radius.

The positive result on the transversal dimension of multi-stage gearboxes is offset by higher mechanical losses, due to the increased number of engaged gear wheels.

It should be noted that in triple stage gearboxes, shown in the picture, the axes of the three shafts do not lie in the same plane, as the scheme seems to show. In a lateral view, the outline of the three shafts should be represented as the vertices of a triangle; this lay-out reduces the transversal dimension of the gearbox. In this case and others, as we will show later, the drawing is represented by turning the plane of the input shaft and of the counter shaft on the plane of the counter shaft and of the output shaft.

Gear trains used in reverse speed are classified separately. The inversion of speed is achieved by using an additional gear. As a matter of fact, in a train of three gears, the output speed has the same direction as the input speed, while the other trains of two gears only have an output speed in the opposite direction; the added gear is usually called *idler*.

The main configurations are reported in Fig. 9.2.

In scheme a, an added countershaft shows a sliding idler, which can match two close gears that are not in contact, as, for example, the input gear of the first speed and the output gear of the second speed. It should be noted that, in this scheme, the drawing does not preserve the actual dimension of the parts.

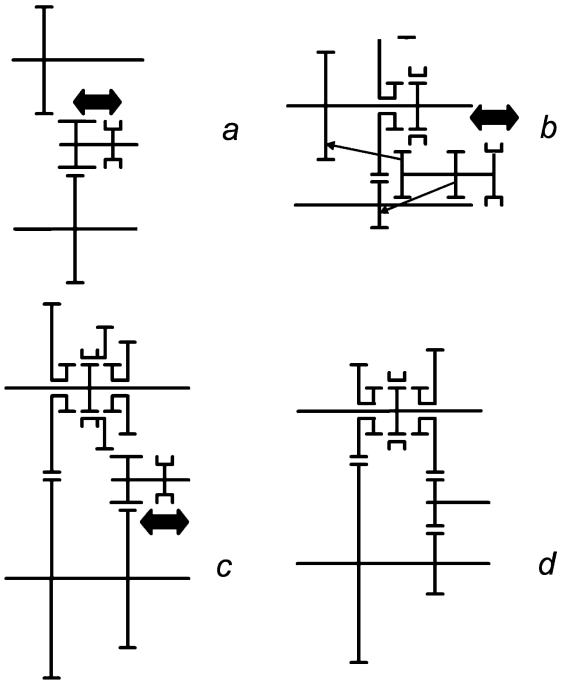


FIGURE 9.2. Schemes used for reverse speed; such schemes fit every type of gearbox lay-out.

Scheme b shows instead two sliding idlers, rotating together; this arrangement offers additional freedom in obtaining a given transmission ratio. The countershaft is offset from the drawing plane; arrows show the gear wheels that match when the reverse speed is engaged.

Scheme c is similar to a in relation to the idler; it pairs an added specific wheel on the output shaft with a gear wheel cut on the shifting sleeve of the first and second speed, when it is in idle position.

Configuration d shows a dedicated pair of gears, with a fixed idler and a shifting sleeve.

The following are the advantages and disadvantages of the configurations shown in the figure.

- Schemes a, b and c are simpler, but preclude the application of synchronizers (because couples are not always engaged), nor do they allow the use of helical gears (because wheels must be shifted by sliding).

- Scheme d is more complex but can include a synchronizer and can adopt helical gears.

- Schemes a, b and c do not increase gearbox length.

9.2 MECHANICAL EFFICIENCY

The mechanical efficiency of an automotive gear wheel transmission is high compared to other mechanisms performing the same function; indeed, the value of this efficiency should not be neglected when calculating dynamic performance and fuel consumption. The continuous effort of to limit fuel consumption justifies the care of transmission designers in reducing mechanical losses.

Total transmission losses are conveyed up by terms that are both dependent and independent of the processed power; the primary terms are:

- Gearing losses; these are generated by friction between engaging teeth (power dependent) and by the friction of wheels rotating in air and oil (power independent).
- Bearing losses; these are generated by the extension of the contact area of rolling bodies and by their deformation (partly dependent on and partly independent of power) and by their rotation in the air and oil (power independent).
- Sealing losses; they are generated by friction between seals and rotating shafts and are power independent.
- Lubrication losses; these are generated by the lubrication pump, if present, and are power independent.

All these losses depend on the rotational speed of parts in contact and, therefore, on engine speed and selected transmission ratio.

Table 9.1 reports the values of mechanical efficiency to be adopted in calculations considering wide open throttle conditions; these values consider a pair of gearing wheels or a complete transmission with splash lubrication; in the same table we can see also the efficiency of a complete powershift epicycloidal automatic transmission and a steel belt continuously variable transmission. For the two last transmissions, the torque converter must be considered as locked-up.

TABLE 9.1. Mechanical efficiency of different transmission mechanisms.

Mechanism type	Efficiency (%)
Complete manual gearbox with splash lubrication	92–97
Complete automatic transmission (ep. gears)	90–95
Complete automatic gearbox (steel belt; without press. contr.)	70–80
Complete automatic gearbox (steel belt; with press. contr.)	80–86
Pair of cyl. gears	99.0–99.5
Pair of bevel gears	90–93

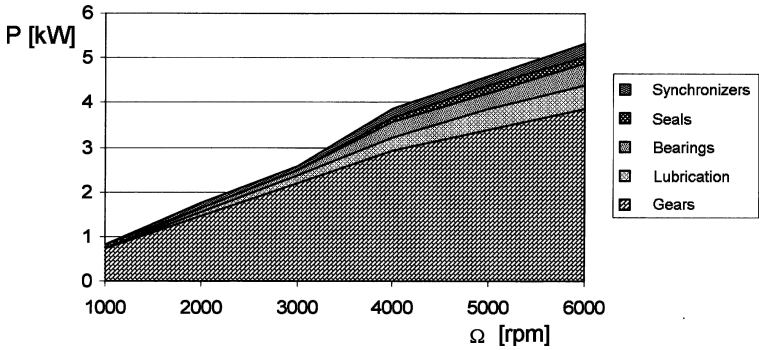


FIGURE 9.3. Contributions to total friction loss of a single stage gearbox designed for 300 Nm as function of input speed.

It is more correct to reference power loss measurement as a function of rotational input speed rather than efficiency. Figure 9.3 shows the example of a double stage transmission, in fourth speed, at maximum power; the different contributions to the total are shown.

This kind of measurement is made by disassembling the gearbox step by step, thus eliminating the related loss.

In the first step all synchronizer rings are removed, leaving the synchronizer hubs only; mechanical losses of non-engaged synchronizers are, therefore, measurable. The loss is due to the relative speed of non-engaged lubricated conical surfaces; the value of this loss depends, obviously, on speed and the selected transmission ratio.

In the second step all rotating seals are removed.

In the third step the lubrication oil is removed, and therefore, the bulk of the lubrication losses is eliminated; some oil must remain in order to leave the contact between teeth unaffected.

By removing those gear wheels not involved in power transmission, their mechanical losses are now measurable.

The rest of the loss is due to bearings; the previous removal of parts can affect this value.

A more exhaustive approach consists in measuring the complete efficiency map; the efficiency can be represented as the third coordinate of a surface, where the other two coordinates are input speed and engine torque. Efficiency calculations can be made by comparing input and output torque of a working transmission.

Such map can show how efficiency reaches an almost constant value at a modest value of the input torque; it must not be forgotten that standard fuel consumption evaluation cycles involve quite modest values of torque and therefore imply values of transmission efficiency that are changing with torque.

Figure 9.4 shows a qualitative cross section of the aforesaid map, cut at constant engine speed. It should be noted that efficiency is also zero at input

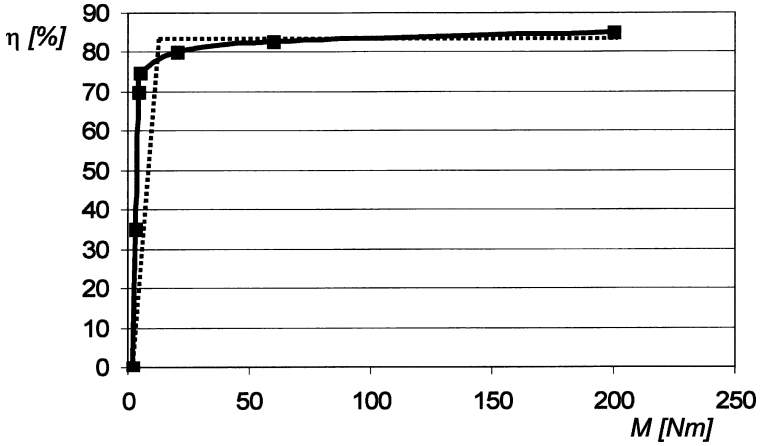


FIGURE 9.4. Mechanical efficiency map, as a function of input torque at constant engine speed; the dotted line represents a reasonable approximation of this curve, to be used on mathematical models for the prediction of performance and fuel consumption.

torque values slightly greater than zero; as a matter of fact, friction implies a certain minimum value of input torque, below which motion is impossible.

A good approximation to represent mechanical efficiency can be made using the dotted broken line as an interpolation of the real curve.

9.3 MANUAL AUTOMOBILE GEARBOXES

9.3.1 Adopted schemes

In manual gearboxes, changing speed and engaging and disengaging the clutch are performed by driver force only.

This kind of gearbox is made with helical gears and each speed has a synchronizer; some gearboxes do not use show the synchronizer for reverse speed, particularly those in economy minicars.

We previously discussed a first classification; additional information is the speed number, usually between four and six.

Single stage gearboxes are used in trans-axles; they are applied, with some exceptions, to front wheel driven cars with front engine and rear driven cars with rear engine; this is true with longitudinal and transversal engines.

In all these situations the final drive is included in the gearbox, which is therefore also called transmission.

Countershaft double stage gearboxes are used in conventionally driven cars, where the engine is mounted longitudinally in the front and the driving axle is the rear axle. If the gearbox is mounted on the rear axle, in order to improve the weight distribution, the final drive could be included in the gearbox.

By multi-stage transmissions, some gear wheels could be used for different speeds. The number of gearing wheels could increase at some speeds; this normally occurs at low speeds, because the less frequent use of these speeds reduces the penalty of lower mechanical efficiency on fuel consumption.

Cost and weight increases are justified by transmission length reduction, sometimes necessary on transversal engines with large displacement and more than four cylinders.

In all these gearboxes synchronizers are coupled to adjacent speeds (e.g.: first with second, third with fourth, etc.) in order to reduce overall length and to shift the two gears with the same selector rod.

We define as the *selection plane* of a shift stick (almost parallel to the xz coordinate body reference system plane for shift lever on vehicle floor) the plane on which the lever knob must move in order to select two close speed pairs. For instance, for a manual gearbox following many existing schemes, first, second, third, fourth and fifth speed are organized on three different selection planes; the reverse speed can have a dedicated plane or share its plane with the fifth speed.

Figure 9.5 shows a typical example of a five speed single stage gearbox. The first speed wheels are close to a bearing, in order to limit shaft deflection.

In this gearbox the total number of tooth wheels pairs is the same as for the double stage transmission shown in Fig. 9.6.

While in the first gearbox there are only two gearing wheels for each speed, in the second there are three gearing wheels for the first four speeds and none

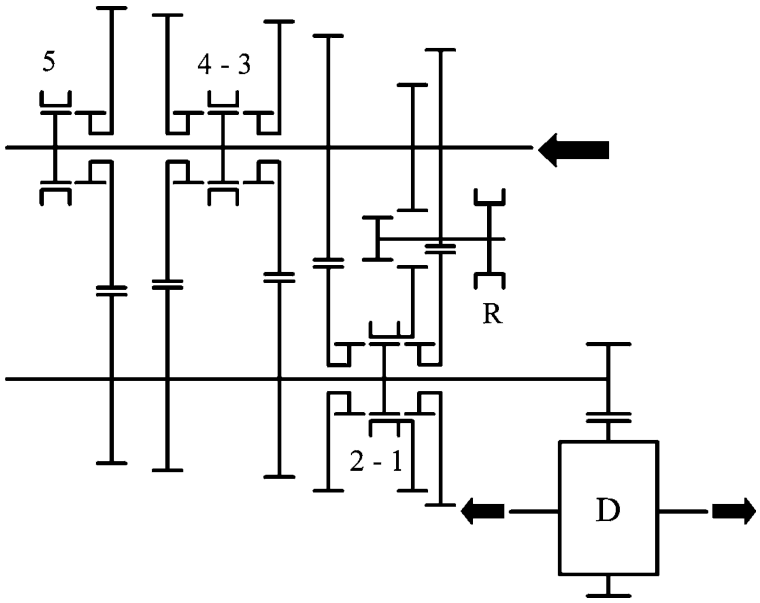


FIGURE 9.5. Scheme for a five speed single stage transmission, suitable for front wheel drive with transversal engine.

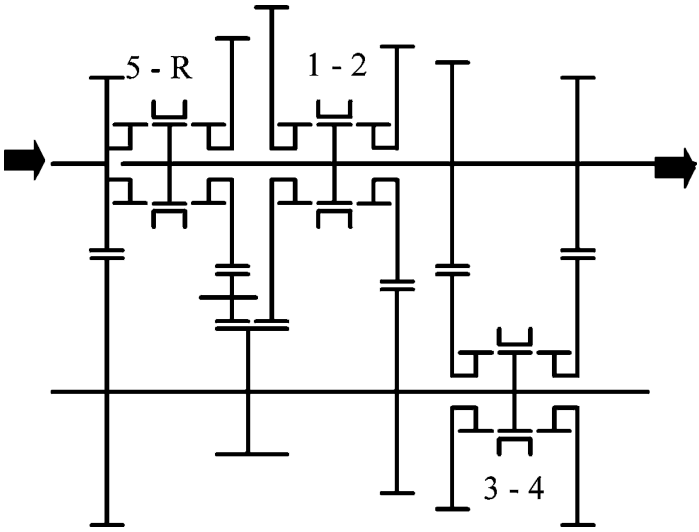


FIGURE 9.6. Scheme of an on-line double stage gearbox for a conventional lay-out.

for the fifth. This property is produced by the presence of the so called *constant gear* wheels (the first gear pair at the left) that move the input wheels of the first four speeds; the fifth speed is a *direct drive* because the two parts of the upper shaft are joined together.

The single stage gearbox in Fig. 9.5 shows the fifth speed wheel pair positioned beyond the bearing, witness to the upgrading of an existing four speed transmission; in this case the fifth speed has a dedicated selection plane.

The double stage gearbox in Fig. 9.7 is organized in a completely different way but also shows the first speed pair of wheels close to the bearing. The direct drive is dedicated to the highest speed; the fifth speed shows a dedicated selection plane.

Six speed double stage gearboxes do not show conceptual changes in comparison with the previous examples; synchronizers are organized to leave first and second, third and fourth, fifth and sixth speeds on the same selection plane.

As already seen, the multistage configuration shown in Fig. 9.7 allows a reasonable reduction of the length of the gearbox. In this scheme, only first and second speeds benefit from the second countershaft; power enters the countershaft through a constant gear pair of wheels and flows to the output shaft at a reduced speed. Third, fourth and fifth speed have a single stage arrangement. Reverse speed is obtained with a conventional idling wheel.

9.3.2 Practical examples

Four speed gearboxes represented the most widely distributed solution in Europe until the 1970s, with some economy cars having only three speeds.

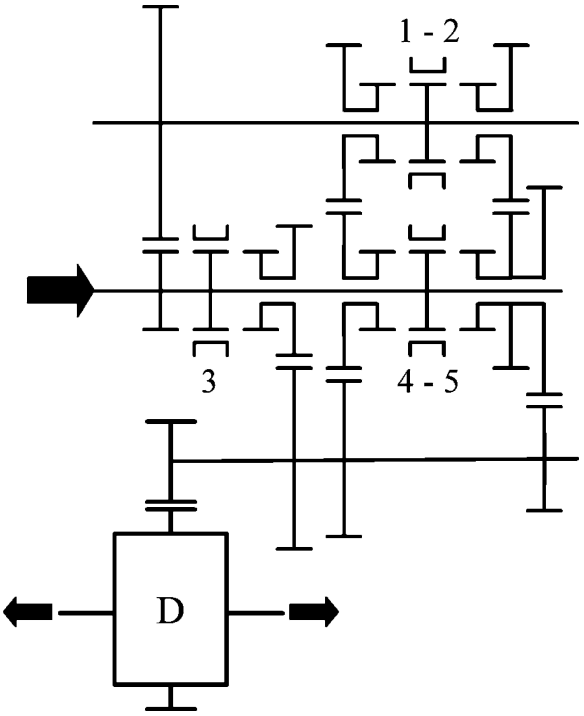


FIGURE 9.7. Scheme of a triple stage five speed gearbox, suitable for front wheel driven car with transversal engine.

With the increase in installed power, the improvement in aerodynamic performance and increasing attention to fuel consumption, it became necessary to increase the transmission ratio of the last speed, having the first speed remain at the same values; as a matter of fact car weight continued to increase and engine minimum speed did not change significantly.

To achieve satisfactory performance all manufacturers developed five speed gearboxes; this solution is now standard, but many examples of six speed gearboxes are available on the market, not limited to sports cars.

Figure 9.8 shows an example of a six speed double stage transmission with the fifth in direct drive; here the first and second pair of wheels are close to the bearing.

This rule is not generally accepted; on one hand having the most stressed pairs of wheels close to the bearing allows a shaft weight containment. On the other hand, having the most frequently used pairs of wheels close to the bearing reduces the noise due to shaft deflection.

Synchronizers of fourth and third speed are mounted on the countershaft; this lay-out reduces the work of synchronization, improving shifting quality by an amount proportional to the dimension of the synchronizing rings. Synchronizers of first and second gear on the output shaft are, because of their diameter, larger

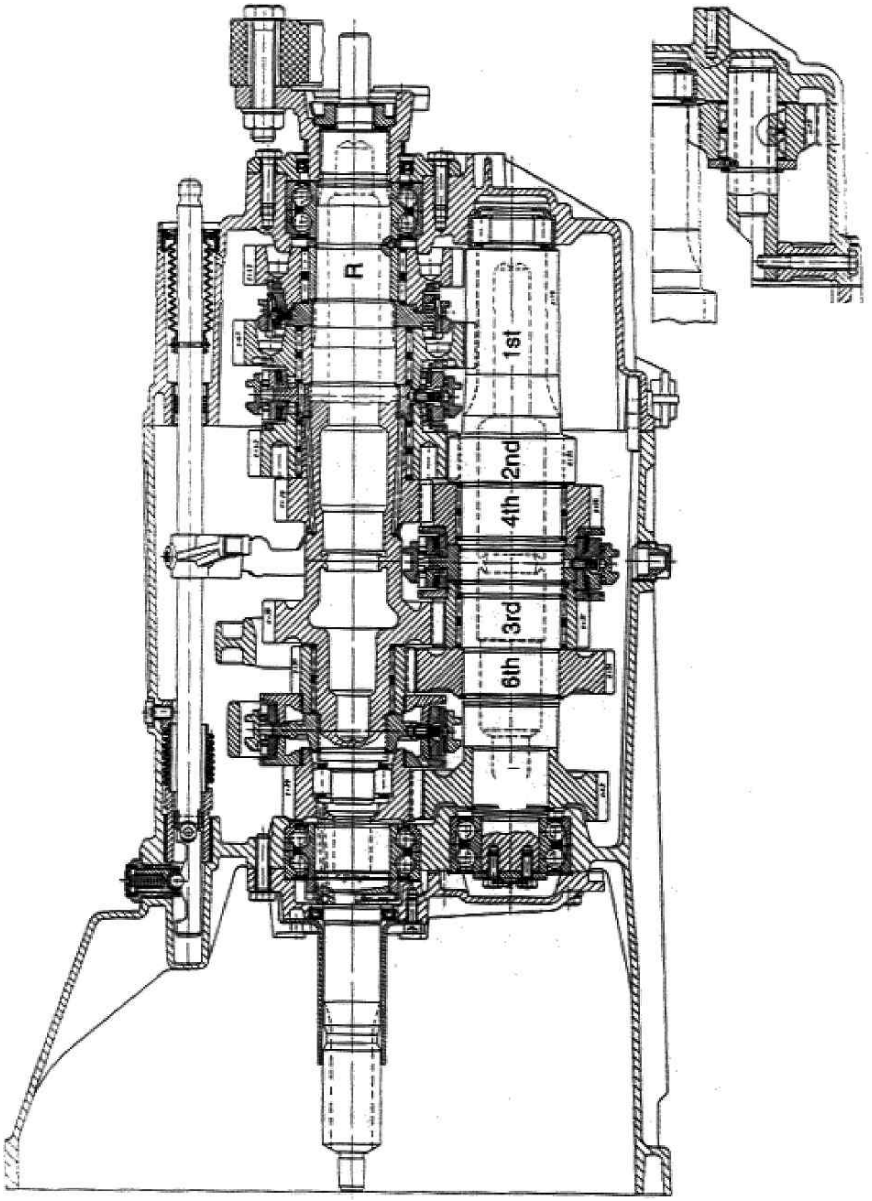


FIGURE 9.8. Double stage six speed gearbox (GETRAG).

than those of the corresponding gear; the penalty of the synchronization work is paid by the adoption of a double ring synchronizer.

Synchronizers on the countershaft offer a further advantage: In idle position the gears are stopped and produce no *rattle*; this subject will be studied later on.

Figure 9.9 introduces the example of a single stage gearbox for a front longitudinal engine. The input upper shaft must jump over the differential, which is set between the engine and the wheels. The increased length of the shafts suggested adopting a hollow section. Because of this length the box is divided into two sections; on the joint between the two sections of the box additional bearings are provided to reduce the shaft deflection.

The input shaft features a ball bearing close to the engine and three other needle bearings that manage solely the radial loads. The output shaft has two tapered roller bearings on the differential side and a roller bearing on the opposite side. This choice is justified by the relevant axial thrust emerging from the bevel gears.

The first and second speed synchronizers are on the output shaft and feature a double ring.

The reverse speed gears are placed immediately after the joint (the idler gear is not visible) and have a synchronized shift. Remaining synchronizers are set in the second section of the box on the input shaft. The output shaft ends with the bevel pinion, a part of the final ratio.

It should be noted that the gears of the first, second and reverse speeds are directly cut on the input shaft, in order to reduce overall dimensions.

Most contemporary cars use a front wheel drive with transversal engine; the number of gearboxes with integral helical final ratio is, therefore, dominant.

In these gearboxes geared pairs are mounted from the first to the last speed, starting from the engine side. An example of this architecture is given in Fig. 9.10.

Like many other transmissions created with only four speeds, it shows the fifth speed segregated outside of the aluminium box and enclosed by a thin steel sheet cover; this placement is to limit the transverse dimension of the power train, in the area where there is potential interference with the left wheel in the completely steered position.

This solution is questionable as far as the total length is concerned but shows some advantage in the reduction of the span between the bearings. Each bearing is of the ball type; on the side opposite to the engine the external ring of the bearing can move axially, to compensate for thermal differential displacements.

One of the toothed wheels of the reverse speed is cut on the first and second shifting sleeve.

The casing is open on both sides; one of these is the rest of one of the bearings of the final drive. A large cover closes the casing on the engine side and, in the meantime, provides installation for the second bearing of the final drive and the space for the clutch mechanism; it is also used to join the gearbox to the engine.

In this gearbox synchronizers are placed partly on the input shaft and partly on the output shaft.

Figure 9.11 shows a drawing of a more modern six speed gearbox, in which it was possible to install all the gears in a conventional single stage arrangement, thanks to the moderate value of the rated torque.

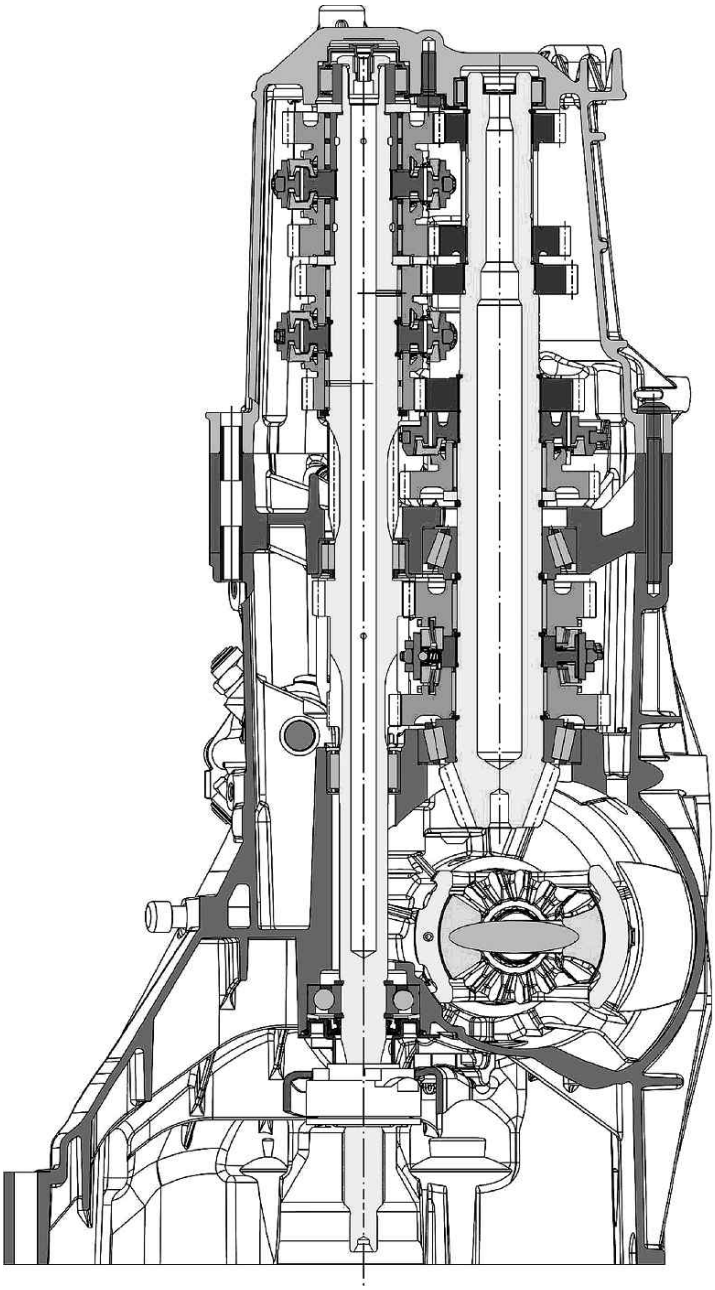


FIGURE 9.9. Single stage six speed gearbox for longitudinal front engine (Audi).

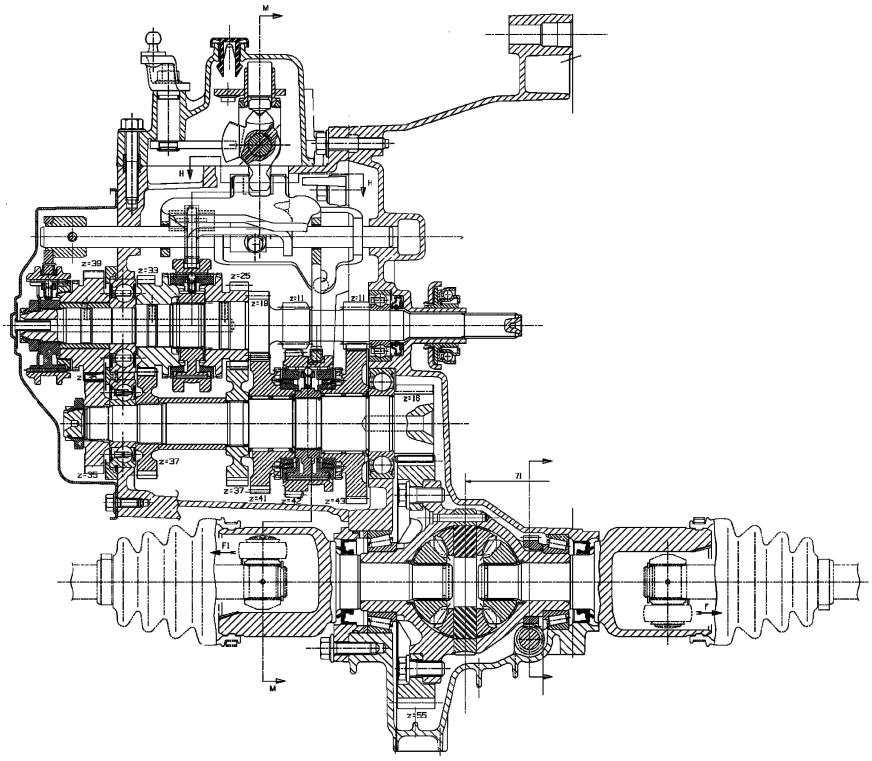


FIGURE 9.10. Five speed transmission for a transversal front engine (FIAT).

Gears are arranged from the first to the sixth, starting from the engine side; as we have already said this arrangement is demanded by the objective of minimizing shaft deflection. Only the synchronizers of first and second speed found no place on the input shaft; they are of the double ring type, as for the first speed.

The reverse speed is synchronized and benefits of a countershaft not shown in this drawing.

9.4 MANUAL GEARBOXES FOR INDUSTRIAL VEHICLES

9.4.1 Lay-out schemes

The gearboxes we are going to examine in this section are suitable for vehicles of more than about 4 t of total weight; lighter vehicles, usually called commercial vehicles, adopt gearboxes that are derived from automobile production, as noted in the previous section.

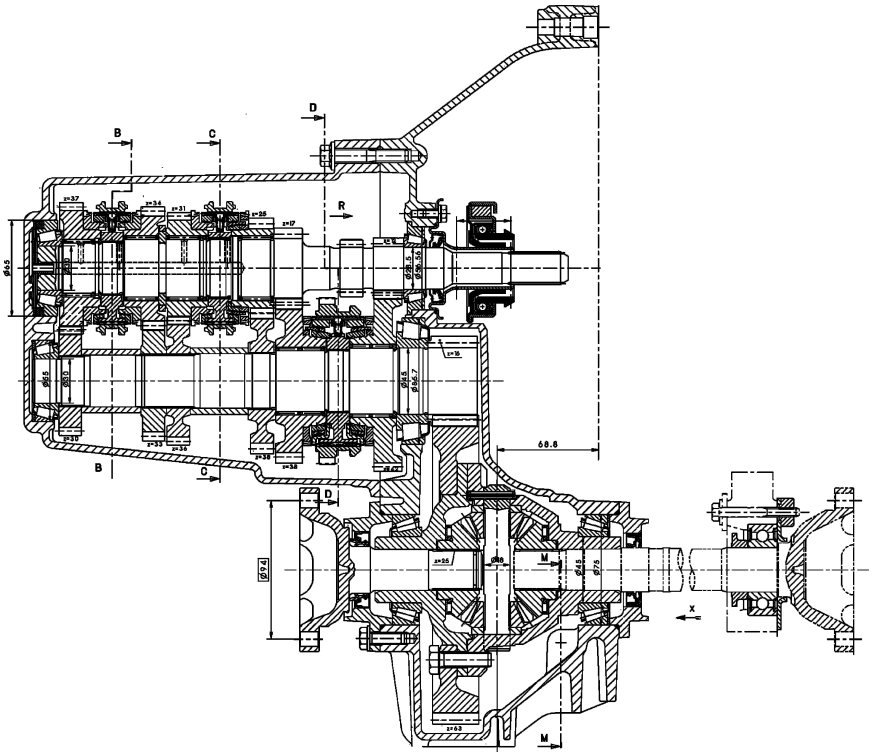


FIGURE 9.11. Six speed transmission for a transversal front engine (FIAT).

Gearboxes used in industrial vehicles also feature synchronizers; they can be shifted directly, as in a conventional manual transmission, or indirectly with the assistance of servomechanisms. Non-synchronized gearboxes are sometimes used on long haul trucks, because of their robustness. Assisted shifting mechanisms are widespread because of the easy availability of power media. Automatic or semi-automatic transmissions are also used, the first type especially in buses.

For gearboxes with four up to six speeds, the double stage countershaft architecture represents a standard; the scheme is the same as seen before.

The constant gear couple is used for all speeds but the highest. Also notable is that the lowest speed wheels are close to the bearings.

As shown in the drawings of Fig. 9.12, the highest speed can be obtained either in direct drive (scheme b) or with a pair of gears (scheme a); in this last case the direct drive is used for the speed before the last: these architectures are called *direct drive* and *overdrive*.

In the figure, only the last and the first before the last speed are represented.

The choice between the two alternatives can be justified by the different vehicle mission; virtually the same gearbox can be used on different vehicles with different frequently used speeds (a truck and a bus for example).

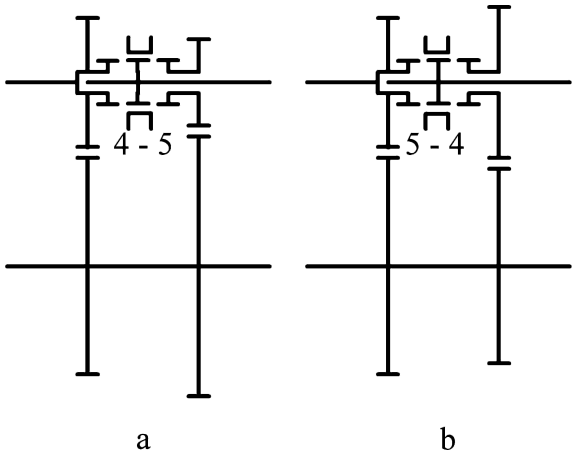


FIGURE 9.12. Alternative constant gear schemes with last or first before the last speed in direct drive.

Sometime the constant gear is set on the output shaft, after the different speed gears; this configuration offers the following advantages:

- Reduction of the work of synchronization, because of the smaller gear dimension at the same torque and total transmission ratio
- Less stress on the input shaft and countershaft

On the other hand, the following disadvantages emerge:

- Bearings rotate faster.
- Constant gear wheels are more highly stressed.

This applies for single range transmissions.

Multiple range transmissions feature, in addition to the main gearbox, other gearboxes that multiply the number of speeds of the main gearbox by the number of their speeds. With this architecture the total number of gear pairs might be reduced, for a given number of speeds, and, sometime the use of the gearshift lever can be simpler.

This arrangement is used when more than six speeds are necessary. A multiple range transmission is therefore made out of a combination of different countershaft gearboxes, single range gearboxes or epicycloidal gearboxes.

Each added element is called a *range changer* if it is conceived as being capable of using the main gearbox speeds in sequence, in two completely non-overlapping series of vehicle speeds; for example, if the main gearbox has four speeds, the first speed in the high range is faster than the fourth speed in the low range.

The element is called a *splitter* if it is intended to create speeds that are intermediate to those of the main gearbox; in this case, for example the third

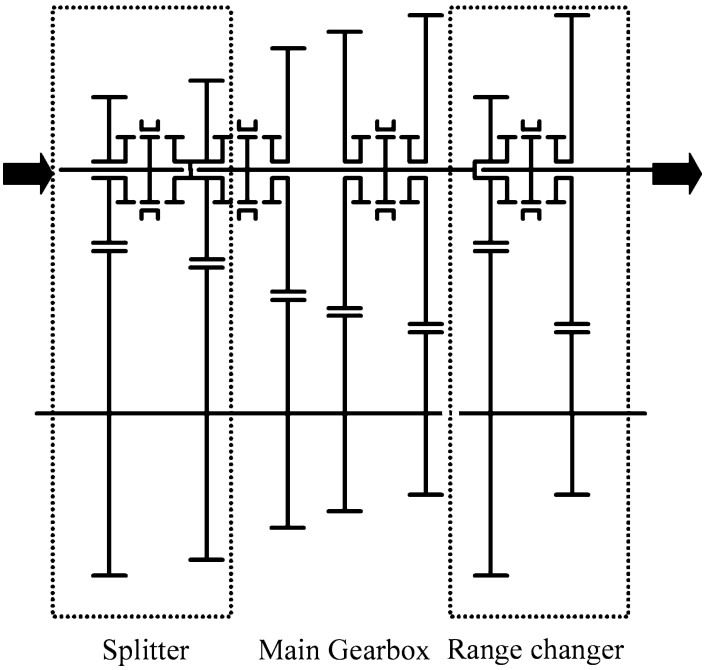


FIGURE 9.13. Scheme of a 16 speed gearbox for industrial vehicles; it is made with a four gear main gearbox, a double speed splitter and a double speed range changer with direct drive.

speed in the high range is faster than the third speed in the low range, but slower than the fourth speed in the low range.

We call the gearbox with the highest number of speeds the main gearbox; the splitter and the range changer will be set in series before and after the main gearbox.

Figure 9.13 shows the scheme of a gearbox featuring a splitter and a range changer. The splitter is made out of a pair of wheels that work as two different constant gears for the main gearbox. The countershaft can therefore be moved at two different speeds, according to the position of the splitter unit. Because the main gearbox has four speeds, this splitter unit can create a total of eight speeds, one of them being in direct drive.

At the output shaft of this assembly, there is a range changer unit made as a two speed double stage gearbox with direct drive; this unit multiplies by two the total number of obtainable speeds. The range changer is qualified by the significant difference between the two obtainable speeds.

The range changer can be made with a countershaft gearbox or an epicycloidal gearbox with direct drive; the advantage in the latter case is the possibility of an easier automatic actuation, by braking some of the elements of the epicycloidal gear.

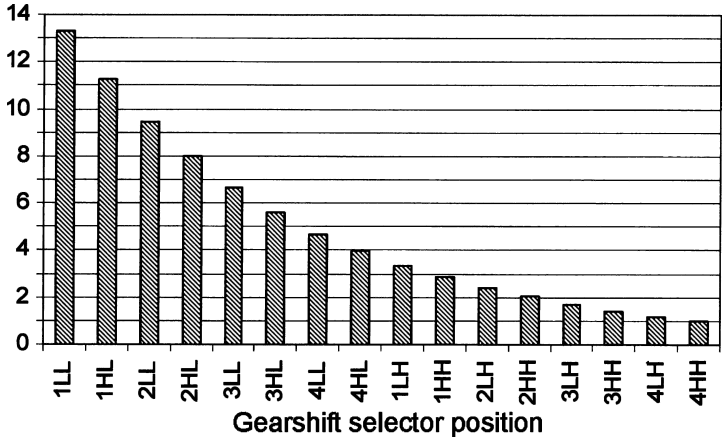


FIGURE 9.14. Transmission ratios obtained with the scheme of transmission shown in Fig. 9.15; speed identification shows the main gearbox speed with the number, the splitter position with the first letter, the range changer position with the second; L stands for low, H stands for high.

It is also possible to place the range changer before the main gearbox and the splitter unit after the main gearbox.

A different way of defining the functions of range change units is to say that the splitter is a gearbox that *compresses* the gear sequence, because it reduces the gap between speeds, while the range changer is a gearbox that *expands* the gear sequence, because it increases the total range of the transmission.

Figure 9.14 explains the concept of compression; the bars represent the ratios obtained in all shifting lever positions. Ratios obtained with the splitter unit in the L position (the first letter in the speed identification, L stands for lower ratio) are interspersed with the ratios obtained with the splitter unit in the H position (H stands for higher ratio, in this case 1:1) and reduce the amplitude of the gear steps of the main gearbox.

The same figure also explains the concept of expansion, showing on the same graph the ratio obtained with the range changer in the H position (second identification letter) and the L position; the gear step between the first in low gear and the first in high gear is as big as the range of the main gearbox, and the total transmission range is widened.

The range changer is therefore seldom used, when driving conditions change suddenly, as, for example, when leaving a normal road for a country road that must be driven more slowly, or when encountering a strong slope with a fully loaded vehicle. The splitter allows the dynamic performance of the vehicle to be improved, making the optimum transmission ratio available to obtain the desired power. The splitter is therefore used frequently. In a fully loaded vehicle, for example, all split ratios can be used in sequence during full throttle acceleration from a standstill.

The range changer and splitter are usually made as modular units that can be mounted at both ends of the main gearbox, or changed with simple covers, in order to satisfy all application needs with limited total production costs.

Generalizing these concepts could suggest building transmissions using additional range changing units arranged in series. These could be conceived as being made only of splitter units with direct drive.

In such a case, with n pairs of tooth wheels it is possible to obtain a total of z transmission ratios, given by the formula:

$$z = 2^{n-1} \ . \tag{9.1}$$

The formula expresses the number of possible states that can be obtained from $n - 1$ pairs of gears; one unit is subtracted because one pair must be a constant gear to move the countershaft.

With four pairs of gears, for example, four speeds can be obtained in a double stage gearbox; while using a cascade of splitters eight different speeds could be obtained. The goal of good shift manoeuvrability and the implications for mechanical losses must not be forgotten, while defining the best architecture.

Figure 9.15 shows the scheme of the 16 speed transmission with splitter and range changer we already described. In this picture are represented the spans of shafts under torque; the dotted line shows where upper input and output shafts are loaded, while the solid line shows when the lower countershaft is loaded. The two lines are joined where a pair of wheels is gearing.

A totally different approach is shown by the double countershaft transmission in Fig. 9.16 (Fuller scheme); the power flow is divided between two countershafts by two constant gears and exits through a single output shaft. This configuration has been conceived with the objective of shortening the gearbox, because it is possible, in this way, to divide the torque on two gear wheels working in parallel. The teeth width can be reduced by about 40% at the same level of rated torque.

On the other hand the transmission is much wider; this choice can represent a favorable compromise for certain vehicles such as road saddle tractors.

In this scheme, after the main gearbox, there is a three speed splitter; one splitter speed is direct drive, one is over-drive, the last is under-drive; the total number of speeds is therefore 12.

It can be noticed that the reverse speed is obtained with the same wheels used for the first speed and with two small idlers. With this arrangement it is also possible to use the splitter in reverse speed.

9.4.2 Practical examples

A practical example of a gearbox for a medium duty truck is shown in Fig. 9.17; in this example a double stage four speed main gearbox is joined to a two speed splitter, offering a total of eight speeds. The splitter unit with its direct drive

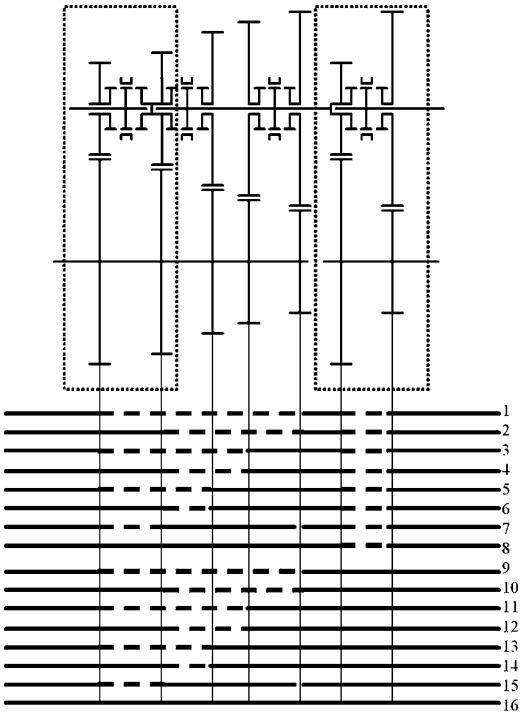


FIGURE 9.15. Power flow scheme in the 16 speeds of a gearbox; lines are dotted when the torque flows through the countershaft. They are solid when the torque flows through the upper shafts (input and output shafts).

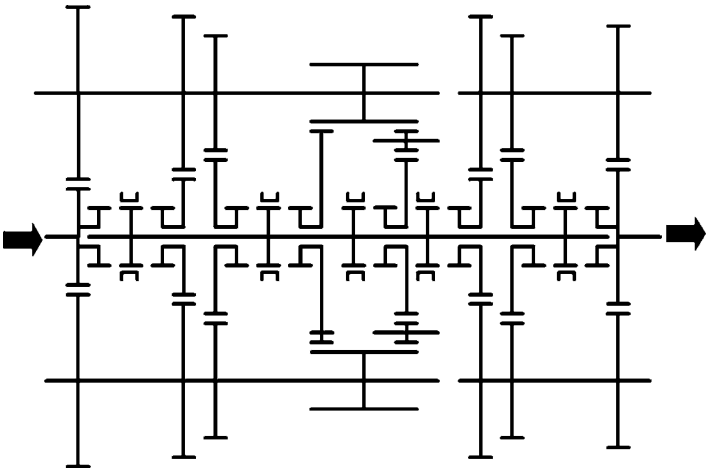


FIGURE 9.16. Scheme of a Fuller gearbox featuring 12 speeds, created from a main four speed double countershaft gearbox and a splitter gearbox with three speeds; one of these last is direct drive. The two countershafts in the main gearbox and in the splitter allow the gear wheel width and therefore the total length of the gearbox to be reduced.

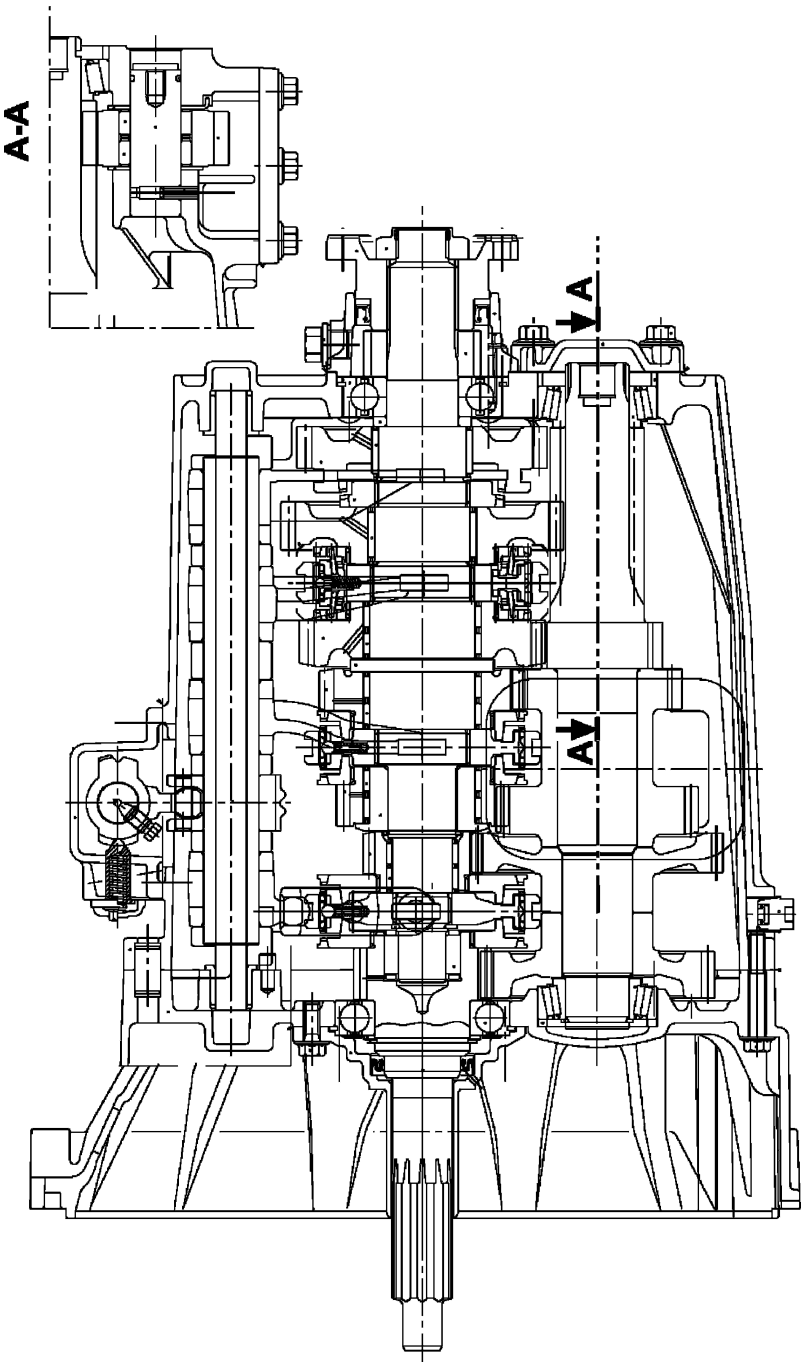


FIGURE 9.17. Truck gearbox with eight speeds; on the lower side there is the reverse idler (IVECO).

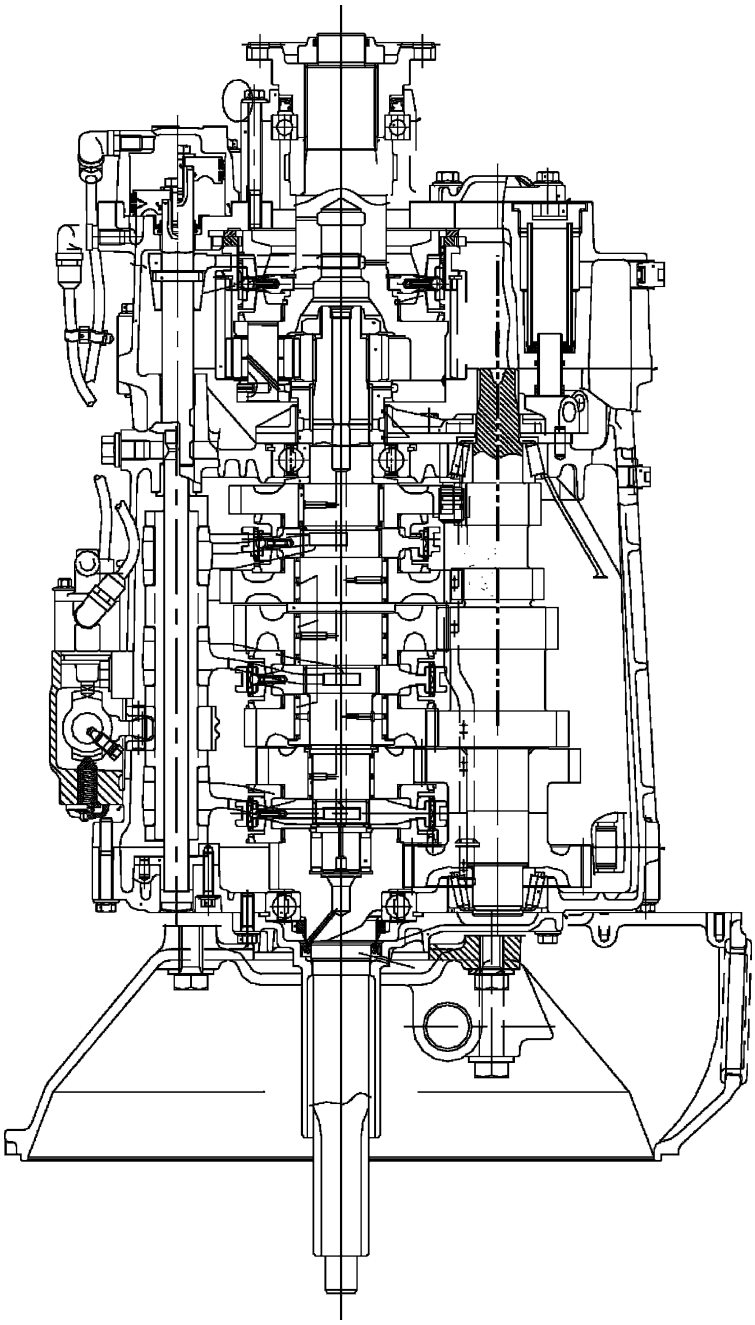


FIGURE 9.18. Truck gearbox with 16 speeds including splitter and range changer (IVECO).

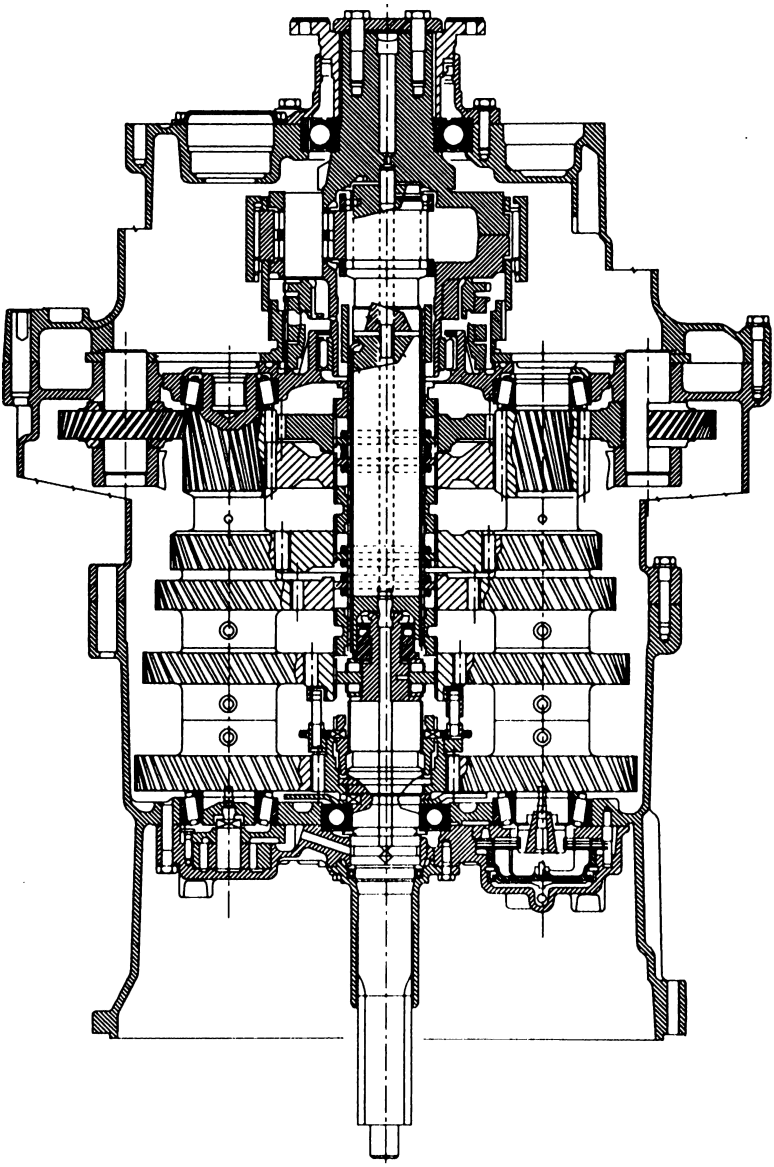


FIGURE 9.19. Fuller gearbox (IVECO).

can obtain the same transmission ratios as the main gearbox, while a reduced speed can obtain transmission ratios that are set between the ratios of the main gearbox.

Section A-A on the lower right side shows a detail of the idler of the reverse speed; the reverse speed is doubled by the splitter.

In the main gearbox the wheels of the first and reverse speeds are close to the rear bearing; the wheels of the following speeds are set in increasing order from the left to the right. The eighth speed is direct drive.

The synchronizers of the first and second speed show a double ring, while the reverse speed has no synchronization. This gearbox can receive a conventional manual shifting mechanism or a power assisted mechanism that can be fully automatic.

A three element gearbox with a total of 16 speeds is shown in Fig. 9.18. A four speed main gearbox (the same as in the previous example) is joined with a two speed splitter and a two speed range changer. The wheels in the main gearbox are ordered in increasing speed from the rear bearing.

The 16th gear is a direct drive. The range changer is made with epicycloidal gears; when the rear shifting sleeve is moved to the left, the epicycloidal gear is blocked and acts like a locked joint.

When the rear shifting sleeve is moved to the right the annulus wheel is blocked with the casing and the carrier speed will be reduced, with the same direction of the input speed.

The reduced speeds stay in a range that is fully separated from the normal speed range; they will be used when a very high torque or a very low speed are needed. The main gearbox shows ball bearings and tapered roller bearings, while the epicycloidal gear train, where the radial thrusts are self-equilibrated, shows only needle bearings and a ball bearing.

The main gearbox countershaft rotates also in idle speed; it shows a spline end that can be used to move auxiliary equipments such as a hydraulic pump useful to operate a tilting loading plane.

A practical example of the Fuller scheme is shown in Fig. 9.19; in this example the gearbox has a total of 16 speeds and is made of a four speed main gearbox, a two speed splitter and a two speed range epicycloidal gear changer.

Notice the two reverse speed idlers. The splitter shift is synchronized, while the main gearbox features dog clutches.

Gear shifts are semi-automatic, manually pre-selected; in this case the gear shift lever does not move the shifting sleeves mechanically, but launches an automatic sequence, where electric valves operate pneumatic actuators. The selection and shift motions do not occur when the lever is moved but when the power is cut off because the accelerator pedal is released or the clutch pedal is depressed.

10
SHIFTING MECHANISMS

10.1 INTERNAL SHIFTING MECHANISMS

Mechanical devices that enable the driver to engage gears in such a way as to obtain the desired transmission ratio are called *shifting mechanisms*. They are called *internal mechanisms* if they are contained in the gearbox casing; they are called *external* when they are mounted partly on the outside of the gearbox casing, partly on the vehicle body; they connect the gearshift lever with the internal shifting mechanisms.

Shifting mechanisms perform an import function, because they are responsible for the response, the driver receives from the shift stick, and therefore for the easy use of the gearbox itself. An ideal response should be positive and precise.

From an ergonomic standpoint, friction and clearance between moving parts of the shifting mechanisms can increase shifting time and make the identification of the engagement positions of the shift stick more difficult. These positions, in fact, are not shown by any device but are learned by the driver through the 'feel' of the shift stick.

Internal mechanisms include selector bars and selector forks; these move the shifting sleeves and synchronizers, as will be explained in a following chapter.

10.1.1 Plunger interlocking device

In order to understand how an internal mechanism works, we will refer to one of the gearboxes seen in the previous chapter; we will not consider, the synchronizer

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

at this time, because its presence does not change the nature of the shifting manoeuvre.

As we have seen, shifting sleeves engage, through an internal spline, with both a hub on the gearbox shaft and another hub on the idle gear wheel. Idle wheels are kept in the correct axial position by spacers and diameter transitions on the shafts and are free to rotate; a gear wheel is locked on its shaft by shifting the sleeve on its hub.

Each sleeve has three discernible positions; with reference to the sleeve on the lower right of Fig. 9.8 for the first and second speed:

- It is in neutral position when engaged with no wheel hubs, as shown in the figure.
- If shifted to the left, it locks the second speed gear to the output shaft and applies the second gear ratio.
- If shifted to the right, it locks the first speed gear to the output shaft and applies the first gear ratio.

The gearbox has four different *sleeves*: one for the first and second speeds, one for the third and fourth speeds, one for the fifth speed and the last for the reverse speed. It should be noticed that in this gearbox the reverse speed is made by shifting an idler, not shown in the figure; when the reverse speed is engaged the idler meshes with the toothed sleeve of second and first (in neutral) and with a specific wheel on the input shaft.

A typical mechanism suitable for this task is shown in Fig. 10.1. It is made by three selector bars (4, 5, 6), each with a fork that engages with a groove on the shifting sleeve; by moving one bar at a time, the desired speed can be shifted. Each bar is moved by a gate 8, existing on each *fork*.

A *finger* lever 3 on a shaft 2 can either rotate around the shaft center line or shift along this centre line. When shifted it selects the correct selector bar; when rotated, in one of the three gates, it shifts the desired speed. The finger matches, with a suitable low clearance, to the gates on the forks.

It should be noted, in this gearbox, that the reverse fork slides on the selector bar of the fifth speed but features its own gate. Acting on levers 1 and 7, shifting and selection motions can be accomplished.

When the gearbox is in neutral, all gates are aligned and the finger is set in the gate of the third and fourth fork.

In order to achieve a proper operation, there should be auxiliary devices suitable for:

- Maintaining the mechanisms in the selected position, even if the driver abandons the shift stick
- Ensuring that more than one bar cannot be moved in the same time, with fatal consequence for the gearbox integrity

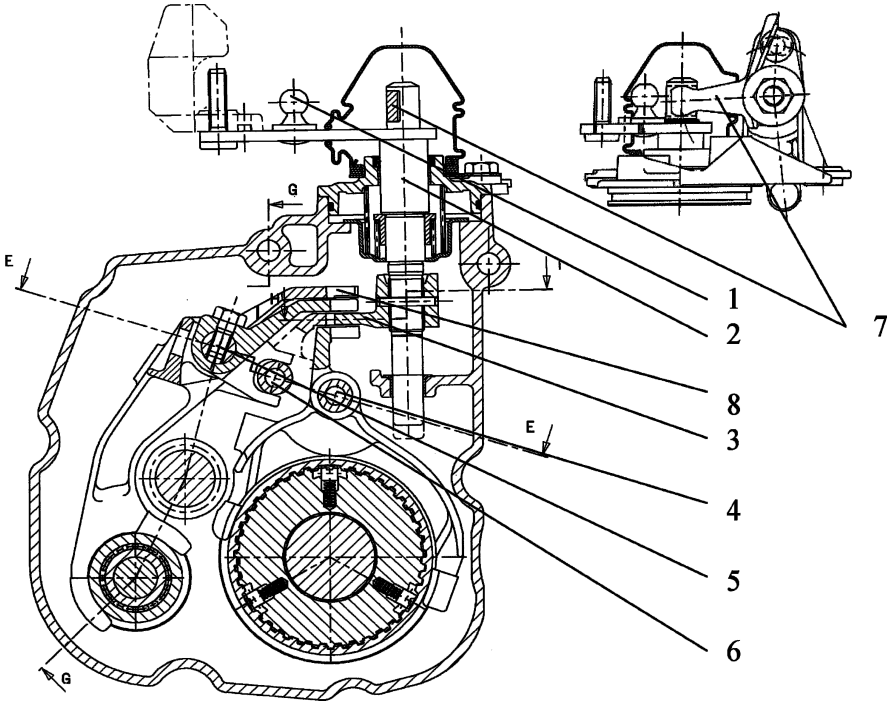


FIGURE 10.1. Cross section of a transversal single stage gearbox on the internal shifting mechanism; 4, 5 and 6 are the shifting bars and forks moving the sleeves; a section of the sleeve on the output shaft, a section of the input shaft (upper left) and a section of the idling shaft of the reverse speed can also be seen (FIAT).

The first function is performed by plungers with spherical ends, matching spherical grooves on the selector bars. The plungers are pushed into their grooves by springs that keep the bar in one of the three discernible positions. Figure 10.2 shows the detail of such grooves (detail 7); the central groove is for the idle position, the other two for shifting the desired speed.

The second function is performed with *plungers* that are similar to those previously discussed; they have their center line on the same plane as the bars' center line (1 and 3 in Fig. 10.2); these plungers are set in two holes in the box, between the bars of first and second speed, and fifth and reverse speed. The bar of third and fourth speed contains a hole, aligned with the others in neutral position, with a needle 2 inside; the length of plungers and needle is designed to allow the motion of one bar at a time.

A spring system must keep the finger in the idle position, as we said, in the third and fourth speed gate; stroke limiters must stop the finger at both ends of the selection stroke, in order to avoid sticking.

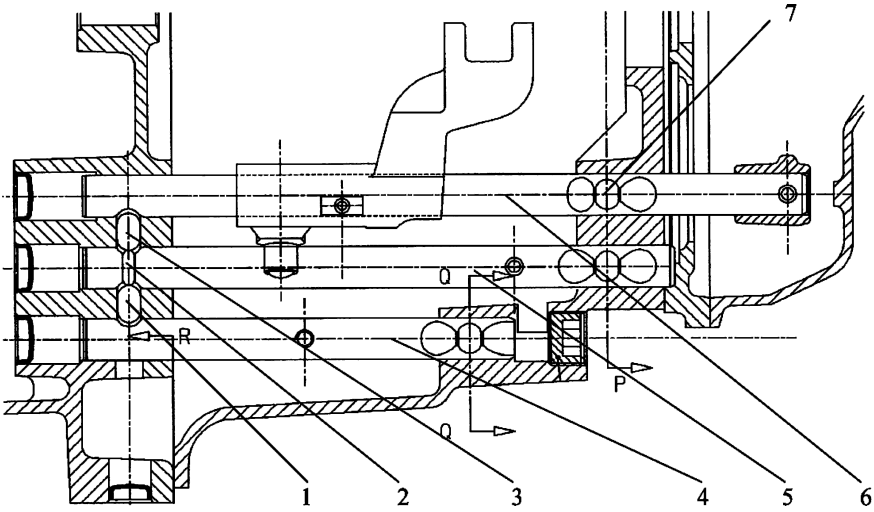


FIGURE 10.2. Detail of the plunger and pin interlocking device, applied on the selector bars, in order to avoid the simultaneous shift of two bars; if one bar is moved the plungers 1 and 3, through the pin 2, keep the remaining bars in idle position (FIAT).

10.1.2 Caliper interlocking device

Clearance between bars and their bearings, shape errors between plungers and their spherical seat, and wear between sliding parts can negatively affect the effectiveness of this mechanism. Even if the contemporary motion of two bars never occurs, parts can stick making the selection difficult or unpleasant.

Figure 10.3 shows an improved solution. There are neither plungers nor grooves. The finger 3 that moves the bars shows in this view on the right of the drawing three spherical seats that can match a ball pressed by a spring (only one is shown). The three corresponding positions fit the neutral position of the three bars.

A U shaped bracket 2, called a *caliper*, with two prongs, matches the cartridge of the ball; the caliper can slide up and down in the neutral position, but cannot rotate. The two prongs match the end of the finger with no clearance.

The caliper allows the motion of one bar at a time, while locking the remaining bars in neutral position.

If the finger is moved in an intermediate position, the caliper would block the entire mechanism in neutral position.

A possible simplification of this mechanism is shown in Fig. 10.4; in this case there is only one bar.

Different from the previous gearbox, where each fork has its bar, in this case each fork, made of stamped steel sheet, slides on the bar of the fifth gear. Each fork shows a gate that can match the finger of the selection mechanism.

These gates are aligned when the forks are in neutral position; devices, not shown in the drawing, avoid any rotation of forks around the bar, leaving only

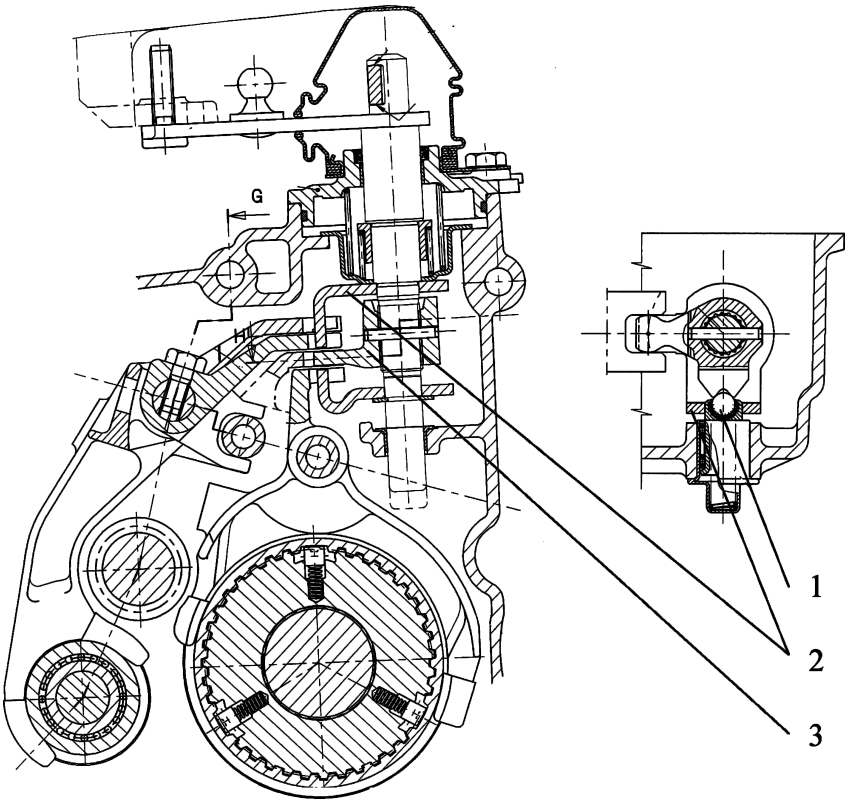


FIGURE 10.3. Cross section of a transversal gearbox with caliper and finger interlocking device; the two prongs of the caliper 2 prevent the finger from moving more than one fork simultaneously; in fact, if the finger could match more than one fork, the caliper prongs would block the selection (FIAT).

the sliding motion possible. In this case the single bar can slide on its bearings and has the double function of moving one fork and guiding the remaining two.

As in the previous gearbox, the caliper shown in Fig. 10.5 moves with the finger only horizontally and avoids the simultaneous selection of two forks.

In order to produce a positive feeling, the bar of the finger slides on ball bearings.

The same care is devoted to the positioning ball of the finger (1 in Fig. 10.4) that has balls both for its rotation and its translation.

10.2 EXTERNAL SHIFTING MECHANISMS

As we previously said, external shifting mechanisms are used to move the internal shifting mechanism as the driver moves the gear shift lever.

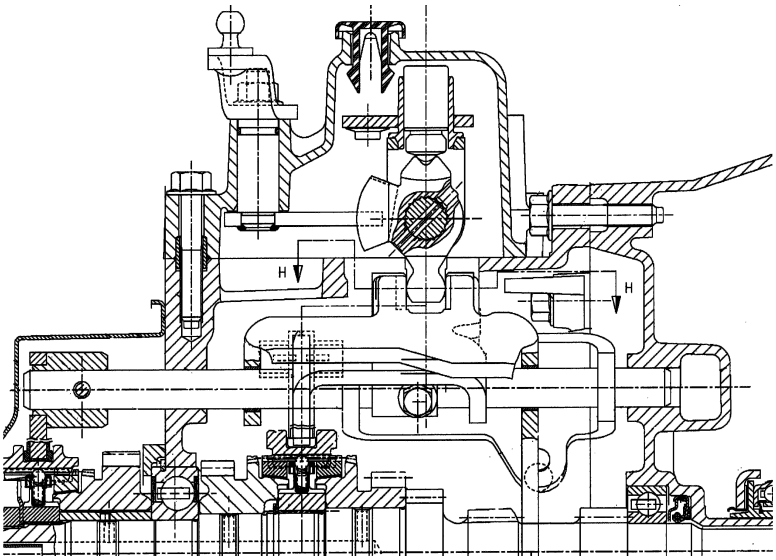


FIGURE 10.4. Cross section of a transversal gearbox; the forks of first and second speeds and of third and fourth speeds can slide on the bar of the fork of the fifth speed (FIAT).

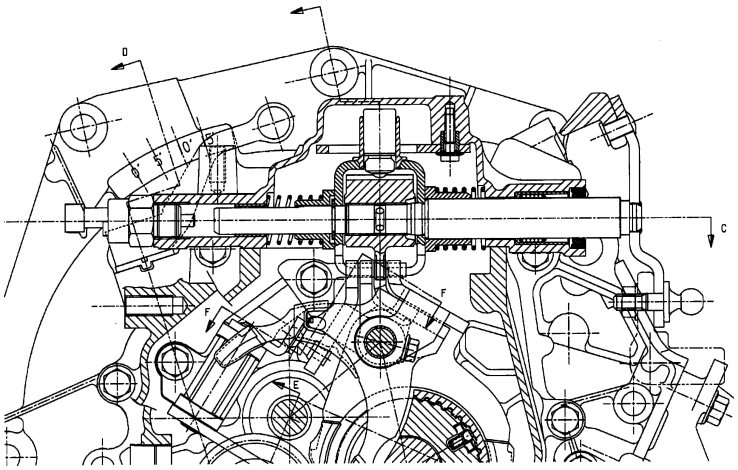


FIGURE 10.5. Cross section of a transversal gearbox; the forks of first and second speeds and of third and fourth speeds can slide on the bar of the fork of the fifth speed (FIAT).

This problem finds an easy solution when the engine is in front of the car in longitudinal position and the traction is applied to the rear axle, because the internal mechanism is close to the natural position of the driver's hand. Selection and engagement motions are coherent with those of the shifting bars and forks.

The engagement positions of the gear shifting lever were historically defined by the simplest solution of the most common architecture.

The solution is much more complicated when engine and gearbox are mounted transversely or when the lever is not mounted on the tunnel, but on the dashboard or on the steering wheel shaft. The tasks to be performed by the shifting mechanisms are two:

- To convey the motion of the lever in a different direction; for example, in a front wheel driven car with transversal engine, the engagement motion in the gearbox is transversal, while the corresponding motion of the shift stick is longitudinal, when the lever is on the tunnel or on the dashboard. It is almost vertical when the lever is on the steering wheel shaft.

- To maintain the engagement position of the shift stick unchanged, with reference to the neutral, even if the power train is moving because of vehicle vertical acceleration or torque variation; this problem is particularly important, because on front wheel driven cars the power train also reacts to the wheel traction force.

- To guarantee a precise and positive feeling on the lever with a limited shifting effort.

We now try to define the attributes given in the previous sentence:

- *Precision* is the capacity to maintain the engagement positions of the shift stick unchanged, in any working condition or, at least, unchanged with reference to the neutral position.

- *Positivity* is the capacity to react to the driver's hand in a consistent way; drivers appreciate limited effort for the selecting motion and the first part of the engagement stroke; higher effort is accepted and expected at the end of the engagement stroke, but must quickly disappear, when the gear has been engaged.

- *Smoothness* is the capacity to limit the variation of the reaction force with reference to an ideal archetype; the reaction force must be not only small but must also show small variation in different manoeuvres; the opposite of smooth is shift stick sticking and slipping.

Stick and slip occurs when two elastic parts are in contact and a relevant friction is present between them: An example is a rubber eraser on drawing paper. The resulting motion between the two elements is characterized by a series of small displacements, with an initial motion due to elastic deformation and a following motion with real slip, which is interrupted when adhesion is again established.

10.2.1 Bar mechanisms

A bar mechanism is a mechanical system made of bars articulated by means of spherical heads. Some bars in the longitudinal direction replicate the motion of the lever in a position close to the engine; other bars in the transversal direction connect the end of these bars with the internal shifting mechanism.

The widest motions of the power train (rotations caused by the reaction torque and vertical oscillations due to uneven road surface) are contained in the xz plane of the vehicle reference system; these motions will not push the transversal bars.

Figure 10.6 shows an example of an application to be matched to a single bar caliper and finger interlocking device gearbox.

The shift stick is connected to the bar 1 for the selection motion, with a pivot with transversal axis; the selection motion will rotate the bar around its axis and will leave the lever free to perform the engagement motion. Bar 1 is mounted longitudinally.

A second bar 2 is linked to the shift stick at a different point, again with a pivot with transversal axis; the engagement motion of the lever shifts this bar longitudinally.

The two motions are decoupled; the end of the bar 1 shows a finger that moves a second bar 3 transversely.

A rocker arm 4 converts the longitudinal motion into transversal motion of the bar 5. Bars 3 and 5 move the internal shifting mechanism in the gearbox.

All articulated heads of bars 1 and 2 and of the rocker arm 4 are mounted on the plates A and B, fixed on the car body.

If bars 3 and 5 have a sufficient length, all power train motions contained in the plane xz will move the lever negligibly. The power train motion in the yz plane is usually small, because the engine suspension in this direction is quite stiff because there are no vibrations to be filtered.

The behavior of this mechanism can be adequate to its mission if the clearance of the pivots is limited and the lubrication of the bearings is acceptable.

This goal is sometime difficult to reach because parts are exposed to dust and splashed water. In addition, the sealing on plate A, which insulates the passenger compartment from dust and noise, where bars cross the firewall, can introduce stick and slip into their motion. These mechanisms have been progressively abandoned in favour of the cable mechanism.

10.2.2 Cable mechanism

Figure 10.7 shows a cable mechanism. In this example the shift stick is mounted on a spherical joint, made on the plastic element 3; two different cable ends are present: 1 for the selection motion, 2 for the engagement motion.

The end 2 is fixed to the lever and copies the engagement longitudinal motion; the end 1 through a rocker arm articulated in 5 also moves longitudinally, but copies the transversal selection motion.

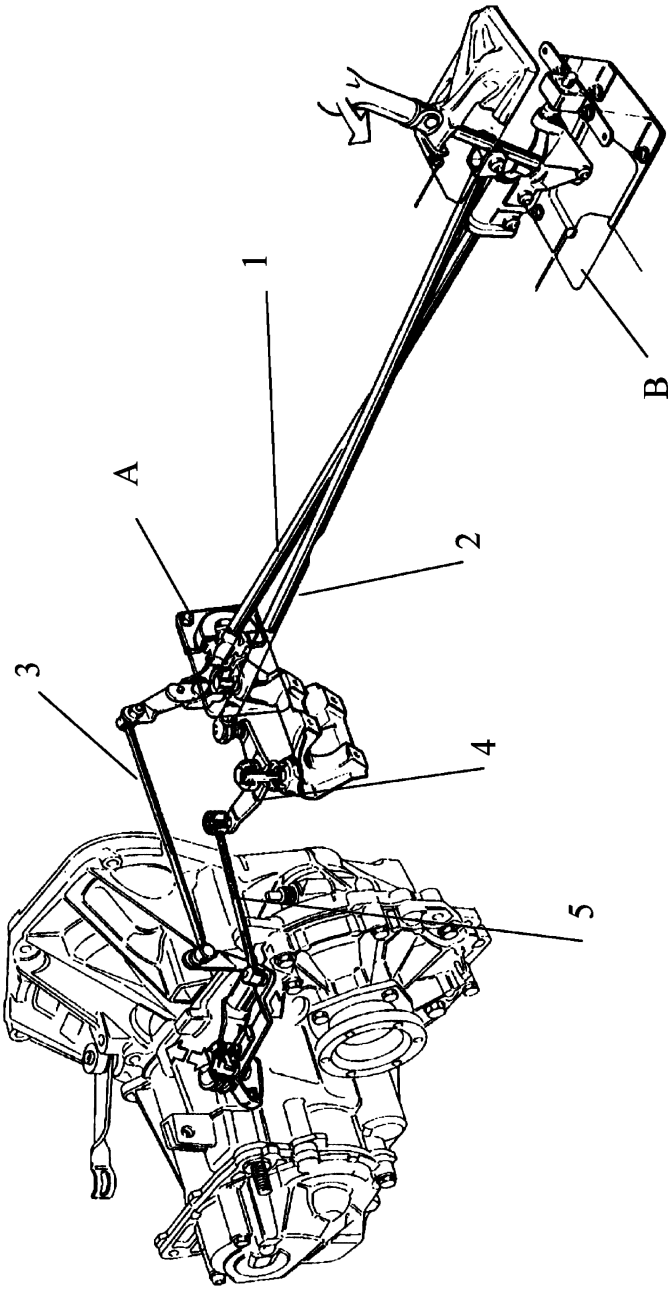


FIGURE 10.6. View of a bar external shifting mechanism; bars 3 and 5 are not affected by the suspension motions of the power train; therefore the suspension motion does not affect bars 1 and 2 and the engagement position of the shift stick (FIAT).

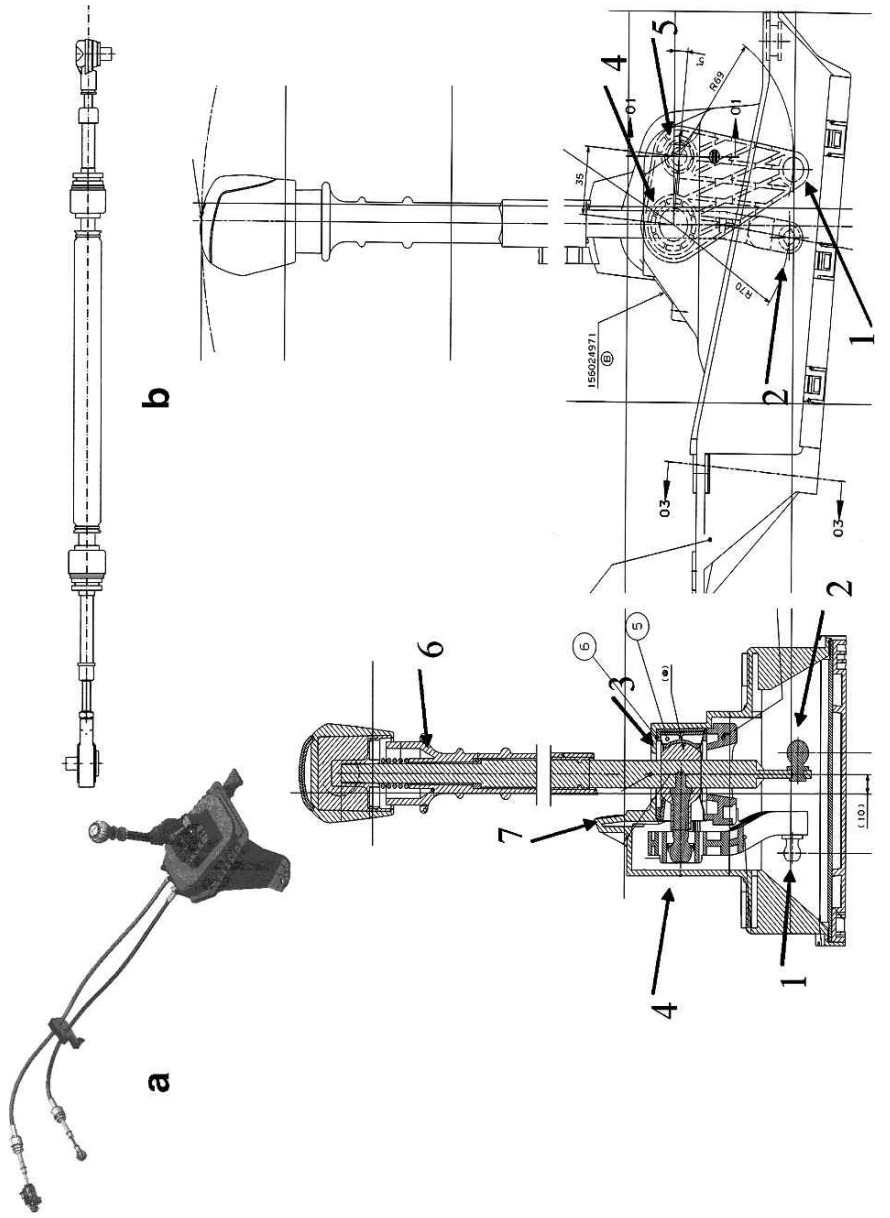


FIGURE 10.7. Detail of a shift stick with cable mechanism; detail A shows the complete lever, while detail B shows the cable itself (FIAT).

It should be noticed that the lever knob has a sliding collar 6 that in its natural position, imposed by a spring, avoids the selection of the reverse speed, because of the interference with the nose 7; reverse speed can be selected only if the collar is voluntarily pulled up.

The internal shifting mechanism is the same as in the previous paragraph. The two basic motions of the lever are transmitted to the internal shifting mechanism by two flexible cables (detail A).

A flexible cable (detail B) is made with a multilayer steel wire sliding inside a sheath. The wires are wound on a cylindrical helix and, thanks to the good lubrication granted by sheath protection are both resistant and flexible.

Cable shows a mechanical resistance higher than that of a wire of the same cross section, but its rigidity is only multiple that of a single wire. The sheath is made with a spirally wound flexible sleeve, inserted in a plastic tube completely sealed from the outside.

The sheath is comparable to a structure which is flexible in bending but stiff in compression; it must be installed with bending radii not smaller than about 100 times the cross section diameter, and with lengths a bit larger than what is necessary, in order to compensate for the power train motion.

If this last can happen without changing the sheath length, the motion of the internal shifting mechanism will replicate the shift stick knob motion very accurately. Smoothness is conditioned by friction between cable and sheath.

Precision and positivity are also affected by mounting stiffness at the sheath ends; the stiffness of these mountings is conditioned by power train vibrations.

In fact, stiff mountings could be quite efficient in transmitting noise and vibration to the driver's hand and to the passenger compartment, but it is important that the brackets sustaining the cable ends are very stiff, so as not to affect negatively the trade-off between manoeuvrability and comfort.

Mechanical efficiency is affected by contacts between cable and sheath; contacts are caused by bending diameter and length. Lubrication conditions and local contacts at the cable ends may also influence mechanical efficiency.

11
START-UP DEVICES

11.1 FRICTION CLUTCH

11.1.1 Clutch functions

Friction is applied in clutches to transmit torque between an input and an output shaft. The clutch is built with three discs; two of these are connected to the engine crankshaft and one, enclosed between the other two, is connected to the gearbox input shaft. Friction between these discs is exploited to transmit and modulate torque. These discs are named respectively the *driving plates* and the *driven plate*.

Driving plates can change their distance from and load applied to the driven plate through a suitable mechanism; the plate can also be disengaged or engaged with an adjustable torque through this mechanism.

In fact a spring system provides the pressure force between driving plates and driven plate, in such a way that friction forces can be established between the discs; the intensity of the pressure determines the value of the transmitted or transmittable torque.

The distinction between *transmittable* and *transmitted torque* refers to two different conditions:

- Clutch completely engaged, without relative motion between the discs
- Clutch partially engaged with relative angular speed between the driving and the driven discs

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

In the first case the transmitted torque is equal to the engine torque and can be at maximum equal to the friction torque between the discs; in the second case the torque is determined by the spring load and is equal to the friction sliding torque between the discs.

The function of the clutch is as follows:

- To transmit the torque from the engine crankshaft to the gearbox input shaft, when there is a difference of speed between the two parts, particularly when the vehicle is stalling or moving backwards
- To connect the two shafts positively, once they are synchronized, in such a way as to transmit all the engine torque
- To separate engine and transmission, when gearbox speed must be changed or the car stopped, without stopping the engine

In addition to these functions, the following functions have been added in recent times:

- To absorb torque pulses caused by the engine polar inertia, in case of misuse of the clutch (main *torsion damper*)
- To control the driveline torsional stiffness, in such a way as to avoid vibration and noise when harmonic frequencies of the engine torque match transmission vibration modes (secondary *torsion damper*)

The springs of the clutch, formerly of the helical type, are now of the *diaphragm* type, because of the advantages thus offered:

- Reduction of power train length, at the same level of torque.
- Reduction of number of parts.
- Easy reduction of weight imbalance, because of geometrical simplicity.
- Simplification of the internal engagement mechanism.
- No effect of wear on maximum transmittable torque.
- Reduction of disengagement load on the clutch pedal.
- Diaphragm springs are insensitive to centrifugal forces; in fact, helical springs with center line parallel to the rotation axis lose load because of their deflection.

Figure 11.1 shows the cut-out view of a diaphragm spring clutch; the engine shaft 1 is flanged to the flywheel 2, which presents one of the active surfaces of the driving plate. Cover 3 is also flanged to the flywheel and puts the second

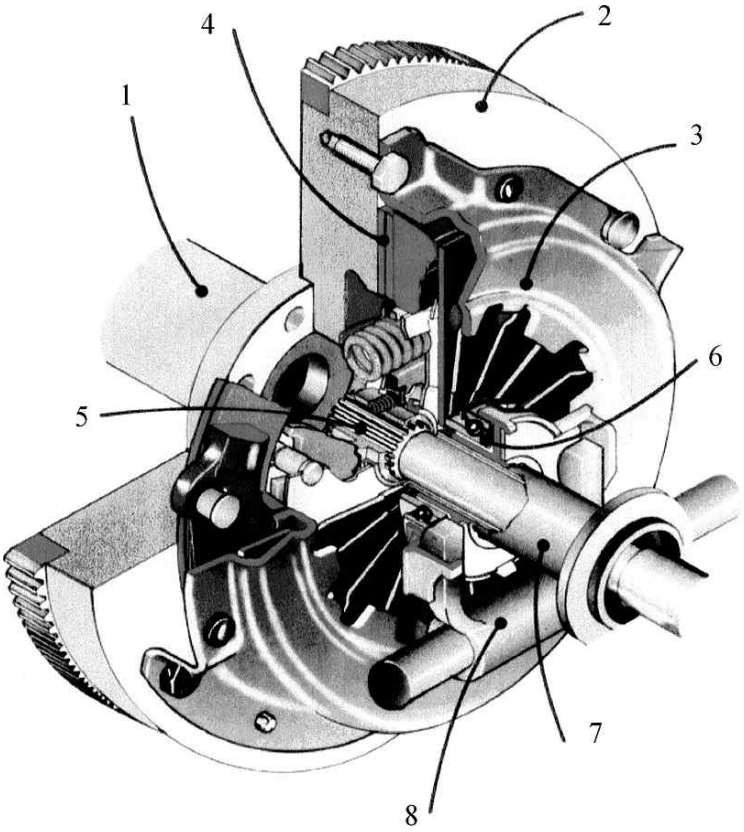


FIGURE 11.1. Cut-out view of a diaphragm spring clutch (Valeo). 1: Engine shaft, 2: engine flywheel and driving plate, 3: cover and internal engagement mechanism, 4: driven plate, 5: gearbox input shaft, 6: thrust bearing, 7: thrust bearing guide, 8: disengagement fork.

driving plate in rotation, pushed by the diaphragm spring. On the driven plate 4 are mounted friction linings; the driven plate has a spline on the gearbox input shaft 5.

The thrust bearing 6 can be axially moved to reduce the reaction pressure of the spring, until the driven plate is disengaged. The bearing can slide on the tube 7 and is moved by the fork 8. The assembly of the moving driving plate, also called the *pressure plate*, of the cover and the related elements utilized for the disengagement form what is called the *disengagement mechanism*.

11.1.2 Disengagement mechanism

Figure 11.2 shows a schematic cross section of a double mass flywheel clutch that will be explained later. If we neglect for the moment the increased complexity of

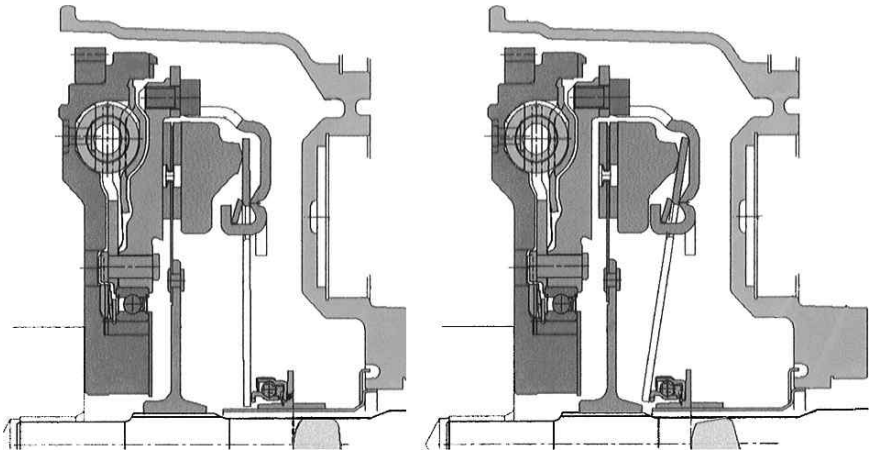


FIGURE 11.2. Cross section of a double mass flywheel clutch, represented at left in engaged condition, at right in disengaged condition (Valeo).

the driving plate, we can consider the clutch as composed of two primary parts: One part is the working surface machined on the engine flywheel. The second is a working surface facing the first, rotating with it, but free to move axially.

The picture represents on the left the engaged clutch, on the right the disengaged clutch. Some important details can be observed: The gearbox input shaft is centred and aligned with the engine crankshaft. The two driving plates should be connected by a device (not shown in this picture) that can transmit the torque but leave the axial motion free; the driven plate is mounted with a spline on the gearbox input shaft, because of the need to adapt its axial position. By moving the thrust bearing axially, the spring diaphragm can be deflected until the force on the pressure plate is brought to zero.

The second element of the driving plate, the pressure plate, shown in Fig. 11.3 with the number 5, is made by a moving disc connected to the stamped steel cover in such a way as to transmit the necessary torque while leaving the disc free to be moved axially.

This goal can be achieved using three flexible straps 6 (3 at 120 deg) that are riveted to the cover and to the pressure plate.

The diaphragm spring 2 is mounted between the pressure plate and the cover and is retained by the ring 4 in the fulcrum 3. All the above elements are part of the disengagement mechanism.

The diaphragm spring is represented in Fig. 11.4; it is made by a tapered steel disc with a number of radial cuts to increase its flexibility; each cut ends with rounded fillets that reduce local stress and are also used as a centering reference.

The spring rests on its larger circumference on the pressure plate and is pressed by the cover through a fulcrum circular area. The spring conical deflection provides the necessary pressure force.

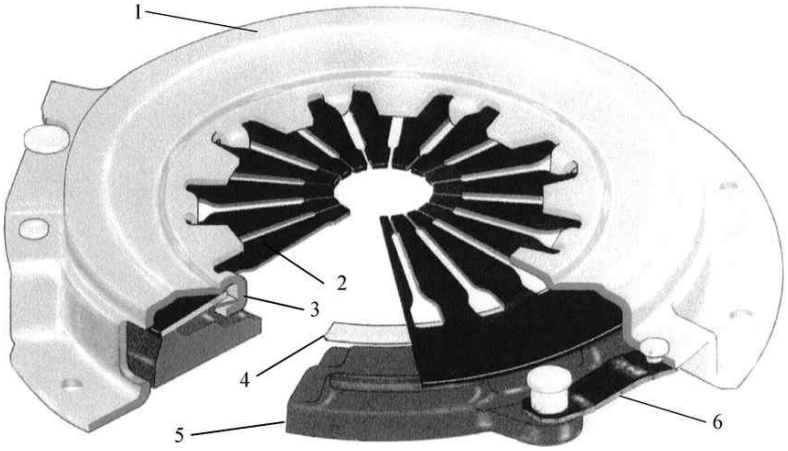


FIGURE 11.3. Cut-out view of a clutch internal disengagement mechanism (Valeo). 1: Cover, 2: diaphragm spring, 3: fulcrum of the spring, 4: ring retainer of the spring, 5: pressure plate, 6: pressure plate strap.

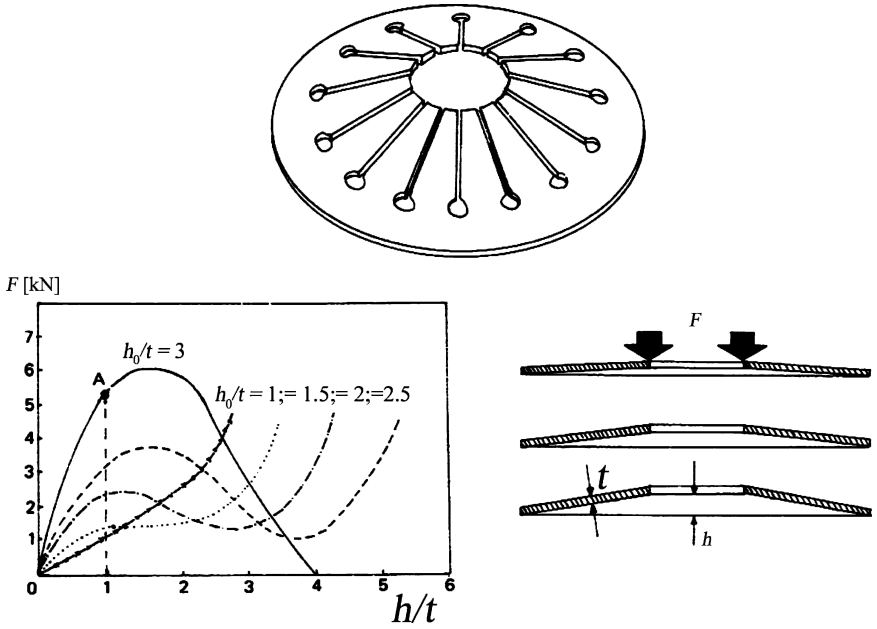


FIGURE 11.4. Diaphragm spring. On the graph at lower left is shown the elastic characteristic of springs as function of the actual height h, divided by the spring thickness t. Curves are parameterized in function of the aspect ratio h_0/t, where h_0 is the unloaded spring height.

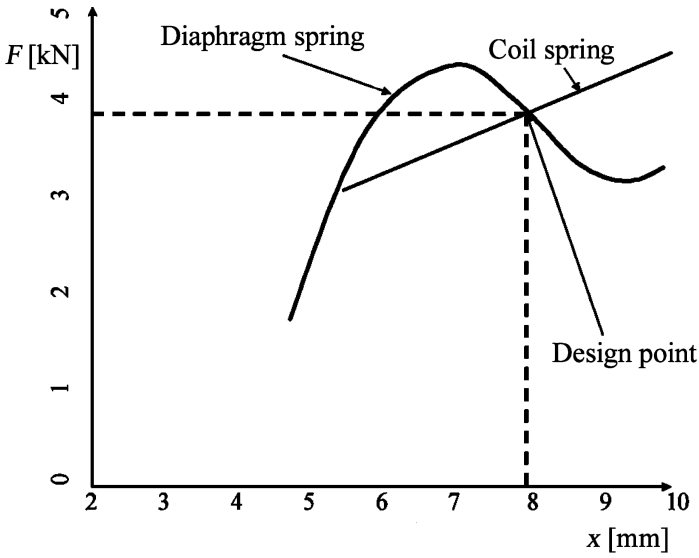


FIGURE 11.5. Qualitative diagram of the pressure plate force F, as a function of the driven plate thickness reduction x caused by wear.

In the same figure is shown the elastic characteristic of the spring. The aspect ratio h_0/t between the height of the conical reference surface (unloaded condition) and the thickness of the diaphragm is used as a parameter of the pressure force-elongation diagram; the diagram is not linear but shows an "S" shape with maximum and minimum values. It is possible to use this feature to avoid the force reduction on the pressure plate, due to the driven plate wear, which is typical on coil springs.

In fact, let us compare the diagrams shown in Fig. 11.5 for a simple coil spring set and a diaphragm spring. Wearing of the friction linings decreases the driven plate thickness during its useful life.

In the case of a coil spring, if we assume that the thickness of the new plate is 8 mm and the maximum design wear is 2 mm, it can be seen that, starting from the new clutch design point, the force on the pressure plate is reduced by spring deformation of about 1 kN.

After a certain period of use pressure can be so much reduced as to cause clutch slippage and driven plate wear-out. On a diaphragm spring of equal specifications, it is possible to maintain the same pressure to the end of the useful life of the lining, without danger of slippage.

There is also a second advantage; if we assume that the disengagement stroke is also equal to 2 mm, the force on the mechanism is increased by about 1 kN with coil springs, while it will be reduced by almost the same value with the diaphragm spring.

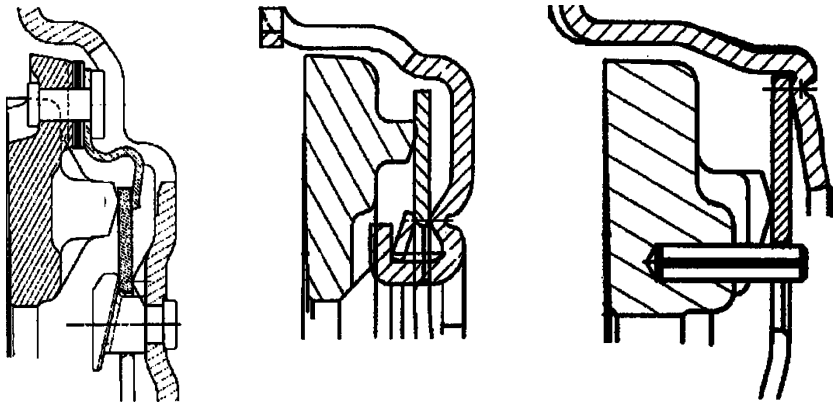


FIGURE 11.6. Internal disengagement mechanisms with pulled disc (the two drawings at left) and pushed disc. Different fulcrum types for the diaphragm spring are also shown.

It should be noticed that as the driven plate wear increases, the spring inner circumference will come closer to the thrust bearing; the bearing must therefore be retracted, in such a way that it is never put in motion when the pedal is released.

The fulcrum between spring and pressure plate can be made using different solutions, shown in Fig. 11.6 for a *pushed spring* (first two figures) and for a *pulled spring* (third figure) version; push and pull refer to the thrust bearing motion during the disengagement stroke.

In the first case the fulcrum circumference is internal to the contact surface of the spring on the disc. The internal leaves of the spring must be pushed against the flywheel to disengage the clutch; the fulcrum can be made with brackets bent on the cover and a ring spacer (second figure) or with riveted pins inserted on the fillets of the spring (first figure).

In the second case the fulcrum circumference is external to the contact surface and the leaves of the spring must be pulled away from the flywheel, in order to disengage the clutch. The advantage of this second arrangement consists of a better use of the available space inside the cover; in fact, at a given outside diameter of the flywheel, the steel cover is less stressed by bending, because the offset between the flange and the contact circumference is reduced.

This advantage is compensated for by a higher cost of the thrust bearing, making this solution suitable for high torque engines of reduced displacement.

11.1.3 Driven plate

Figure 11.7 shows a drawing of a driven plate. Its design is quite complex, especially near the hub where torsional dampers are found.

The plate is built with several steel discs. The first, marked with number 1, is the hub, with a spline fitting the gearbox input shaft. Two other elements

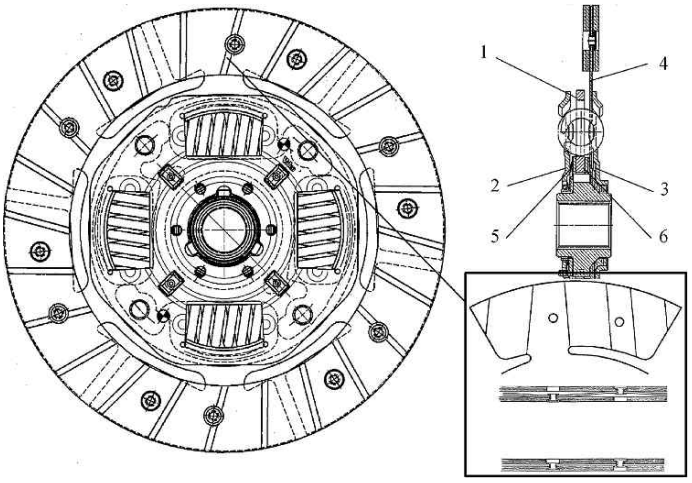


FIGURE 11.7. Driven plate (Valeo). The cross section shows the main and the auxiliary torsion dampers. The detail on the right shows how friction liners are mounted on the disc, in order to make the engagement progressive.

marked 2 and 3 are made with two steel discs riveted with spacers; the element 4 that supports the friction linings is also riveted to the disc 3.

The disc 4 has a particular shape (see also the detail at the lower right). This disc shows a number of separate sectors on which the friction linings are riveted. The sectors are bent off the main disc plane and the rivets set on parallel planes. With this arrangement the friction linings are spaced and the two friction surfaces can displace elastically along the axial direction.

The result is that the pressure force is limited by disc elasticity and clutch engagement is more progressive. In addition, because the thermal gradient on the pressure plate can cause conical deformations, linings can more easily match this shape, maintaining an uniform contact on the surface of the driving discs.

The assembly of discs 2, 3, 4 is free to rotate at a certain angle, with reference to the hub 1.

The three elements 1, 2 and 3 show four rectangular windows, where four coil springs are press fitted. If the torque acting on the disc is less than the total pre-compression moment of the springs, the disc acts as a rigid body; if the torque is higher, disc and hub can assume angular deflection.

Between the rotating elements, there are also two friction discs 5 and 6, pressed through a diaphragm spring; when the hub is moving it is possible to dissipate part of the elastic deformation work of springs and damp torsional vibrations. This assembly is the main *torsion damper*.

In the following Fig. 11.8 a magnification of the hub 1 is shown. We can see that it presents the secondary torsion damper, similar to the previous; the stiffness of this second damper is much lower.

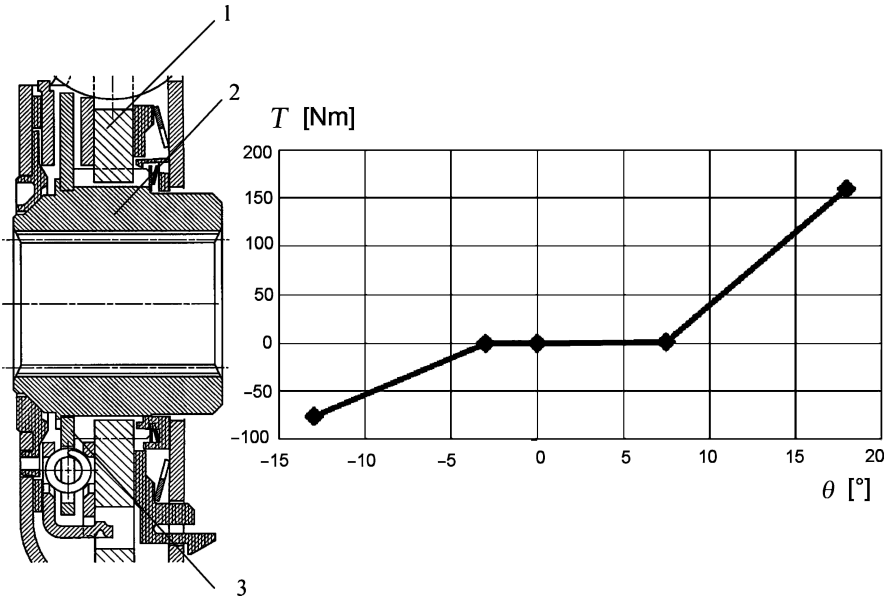


FIGURE 11.8. Magnification of the driven plate hub (Valeo). The diagram on the right shows the angular displacement of the hub θ, as a function of the torque T.

Only the hub 3 is fixed on the input shaft, while the hub 1 can assume significant angular displacements, limited by a number of teeth.

In the same picture is shown a possible elastic characteristic of the hub assembly; the central part of the diagram is very flexible (secondary torsion damper), continuing in two stiffer ends (main torsion damper). The two ends of the diagram can be asymmetric if coil springs in the rectangular windows are double (one inside the other, of different length).

The stiffer damper is finalized to the rattle noise suppression at idle, while the softer damper is finalized to the rattling noise at low speed operation.

In some clutches these functions are not present on the disc, which is very simple, but are fitted on the driving plate, on the engine side; an exploded view of this plate is shown in Fig. 11.9. The flywheel and driving plate are made with two masses free to rotate independently on a ball bearing. Two long coil springs lower the value of the natural frequency of one of the masses; a damping device is also present.

The result that can be obtained with a suitable tuning is that the lowest natural frequencies of the transmission, usually between 800 and 2,000 rpm, can be shifted below the engine idle speed, and over the maximum speed, where the transmission noise cannot be detected by human ears; the results are particularly interesting in high displacement engines.

The total inertia of the system is the same as that of the reference flywheel; there are therefore no effects on engine torque fluctuations.

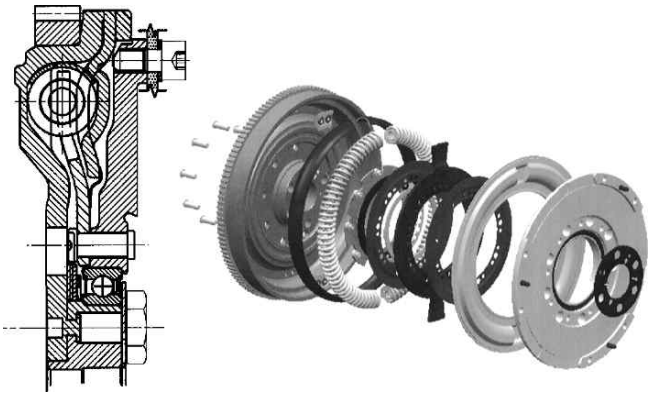


FIGURE 11.9. Cross section and exploded view of a double mass clutch and flywheel. The torsional damper is now included in the driving plate; the driven plate is much simplified (Valeo).

11.1.4 Thrust bearing

The thrust bearing, as we have seen, slides on a tube flanged to the gearbox housing, coaxial to the input shaft (see part 7 in Fig. 11.1); this tube is usually integrated with the gearbox oil seal cover.

The thrust bearing is moved by a finger lever on the gearbox housing and by the clutch pedal through an external mechanism. The bearing on the left side of Fig. 11.10 has a simple shape and acts on the leaves of the diaphragm springs directly.

The external disengagement mechanism can by made with a cable or a hydraulic actuator.

The cable mechanism is similar to that explained for the external shifting mechanisms of the gearbox. As we have said, the cable assembly must be installed with no sharp bends that affect mechanical efficiency negatively; poor mechanical efficiency can increase the disengagement pedal force and introduce a hysteretic behavior that can make pedal return disagreeable or difficult.

The progressive wear of the friction linings on the driven plate must be compensated for with a corresponding adjustment of the thrust bearing position; this adjustment shifts the useful stroke of the pedal closer to the driver's foot, with possible negative effects on ergonomics. To avoid this inconvenience and to keep the stroke positioned correctly, an adjustment nut must be provided on one of the rests of the cable sheath.

Hydraulic circuits are progressively replacing cable mechanisms. A hydraulic cylinder is linked to the pedal and operates a hydraulic actuator through a pipe acting on the disengagement fork of the thrust bearing; it is sometimes possible to integrate this actuator with the clutch housing in such a way as to push directly on the thrust bearing, with a simplification of the mechanism, as shown in Fig. 11.10.

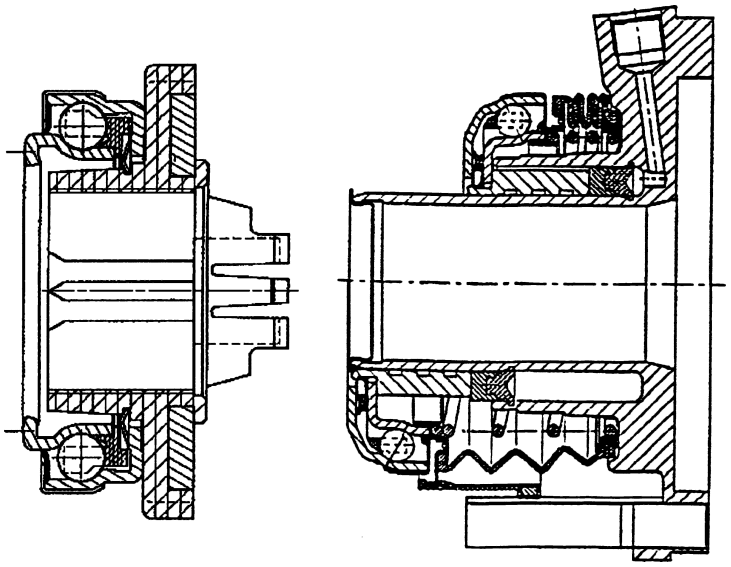


FIGURE 11.10. Conventional thrust bearing (on the left) and bearing with integrated hydraulic actuator (Valeo).

The hydraulic mechanism has the advantage of better mechanical efficiency which is almost independent of piping lay-out; with a simple check valve between the oil reservoir and the operating piston on the pedal side, it is easily possible to obtain an automatic adjustment of the pedal stroke position.

Driven plate wear, also with pedal stroke adjustment, has the inconvenience of increasing the pedal actuation force, because of the elastic characteristics of the diaphragm spring, as shown in Fig. 11.5; in addition, the inner space left inside the clutch cover must take into account the adjustment of bearing stroke caused by wear.

A definitive solution to all these problems is presented by the self-adjusting internal mechanism shown in Fig. 11.11. In this mechanism the diaphragm spring pushes the pressure plate through a spacer 1, made of two steel sheet discs connected with flexible straps; the contact surface with the pressure plate is a helical ramp, so that by rotating the spacer the driven plate thickness decrease can be compensated for.

The spacer is turned by means of teeth 2 that match with the worm gear that is magnified on the right side of the picture; if, during a clutch engagement the driven plate thickness has decreased, the leaf spring 3 contacts the diaphragm spring and rotates the worm gear until the contact between the two springs is lost.

With this device it is possible to keep mechanical clutch characteristics at the design value and to reduce the axial length of the clutch cover.

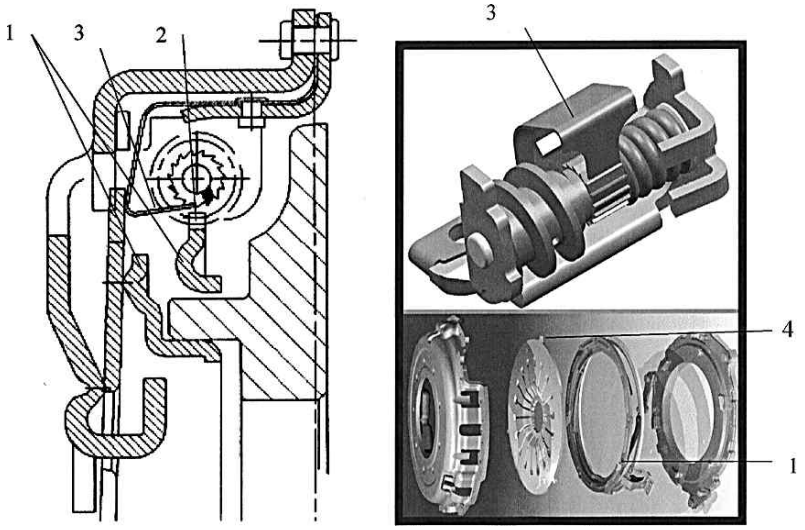


FIGURE 11.11. Self-adjusting mechanisms compensating for the effects of wear on the driven plate thickness (Valeo).

Driven plate bending vibrations are caused by the engine crankshaft. During high speed engine disengagement manoeuvres, they can cause disagreeable noise and pedal vibrations; both are transmitted through the disengagement mechanism.

This fact justifies the application of rubber elastic bearings on sheath ends or a hydraulic capacity on the oil circuit.

In case of severe vibration problems, for example in high speed sport engines, when the noise propagates through other ways, the crown of the flywheel is mounted on its hub with a flexible plate that compensates for crankshaft end fluttering.

11.1.5 Design criteria

When designing a clutch many points must be considered:

- The disengagement force on the pedal should not be too excessive.
- The friction coefficient must be as constant as possible, in a wide range of operating conditions.
- The wear on the working surfaces must be gradual.
- The heat produced during engagement manoeuvres must be dissipated.
- The clutch must operate correctly and with reasonable wear.

The friction coefficient between discs and friction linings ranges between 0.27 and 0.30 at acceptable temperatures.

Taking this value into account the diaphragm spring can be designed in such a way that maximum engine torque can be transmitted with a safety margin of about 15%, with both new and completely worn linings.

It is possible to calculate pedal force with new and worn linings from spring characteristics; the mechanical efficiency of the transmission (between pedal and disengagement fork) must also be taken into account.

Typical values for a sound cable mechanism range between 0.65 and 0.70; for a hydraulic mechanism with piston and disengagement fork, the value can reach 0.80. For an integrated hydraulic actuator the efficiency can be 0.90.

Friction is therefore not negligible in the calculation of the maximum disengagement thrust (which must be kept below 300 N); this value cannot be reduced by working on the transmission ratio between pedal and bearing only, because the pedal stroke must be kept below 150–200 mm; the elastic deformation of the mechanical chain between pedal and thrust bearing must also be considered.

A useful design parameter is the mechanical energy dissipated at start-up; it can be referred to the useful friction surface of the clutch and compared with similar values obtained in successful cases; the same can be said for the temperature at the end of a start-up manoeuvre at maximum slope.

11.2 START-UP DEVICES FOR AUTOMATIC GEARBOXES

Gearbox automation demands devices that can start-up the vehicle smoothly, without the help of driver perception; for this reason, devices different from the clutch were developed, utilizing intermediate non-mechanical energy, such as electric or hydraulic.

Today this issue is not as important, because it has proven possible to automate a conventional friction clutch, designed for pedal actuation, by using an electronically controlled electro-hydraulic actuator, with satisfactory results.

The hydraulic torque converter and the electromagnetic powder clutch, the latter on very few applications, are used with automatic gearboxes, for their intrinsic comfort characteristics; they can only be matched by gearboxes whose speeds can be changed without torque interruption. As a matter of fact, the high rotating masses of these devices will cause a conventional synchronizer to work with an extremely long engagement time.

The powder electromagnetic clutch is made with two ferromagnetic coaxial rings that build the poles and the anchor of a magnetic circuit; the magnetic field can be created by an electric coil on the outside ring, fed with sliding contacts.

The air gap is several millimeters thick; in the air gap a quantity of iron powder with suitable granular dimension is inserted. The magnetic field that builds up when the current is switched on causes the iron grains to align along

the radial field lines, between the two poles; grains exchange friction forces with each other and with the polar surface. Friction brings the two poles to synchronize while exchanging a torque depending on the current intensity.

The electromagnetic clutch is easy to control; on the other hand the long response time and high inertia are not suitable to a synchromesh gearbox. The use of this clutch is justified only by the reduced cost in comparison with hydrodynamic clutches and torque converters.

The torque converter is a particular hydraulic machine that allows two shafts with a continuously variable transmission ratio to be connected. Unlike the friction clutches, the input and the output torque values of the torque converter are not bound to be equal, but they are determined by more complex relationships depending on the transmission ratio.

If we call Ω the angular speed, M the torque and P the power and associate subscripts 1 and 2 to these magnitudes on the torque converter input and output shafts, we define:

- *Speed transmission ratio*:
$$\nu = \frac{\Omega_2}{\Omega_1} \ .$$
Sometimes $1 - \nu$ or *slip* is used, in place of the speed transmission ratio.

- *Torque transmission ratio*:
$$\mu = \frac{M_2}{M_1} \ .$$
Transmission efficiency:.
$$\eta = \frac{P_2}{P_1} \ .$$
Because, by definition:
$$P_1 = M_1\Omega_1 \text{ and } P_2 = M_2\Omega_2 \ ,$$
as a consequence:
$$\eta = \nu\mu. \tag{11.1}$$

Therefore, also neglecting mechanical friction, mechanical efficiency can be less than one in a torque converter; transmission ratios are governed directly by the internal geometry of the machine and cannot be easily changed by means of external devices.

In order to understand the converter operation it would be useful to think about a transmission made by a hydraulic pump and a hydraulic turbine connected to each other with suitable piping.

The pump rotates with the engine and moves a certain flow of oil from a suction pipe to an outlet pipe; the pipes are also connected to a turbine that is mechanically linked to the gearbox input shaft.

A transmission like this has been used on ships to connect engines with propellers.

Unlike mechanical transmissions, there is no positive link between the two shafts. The transmission ratio is determined by the inertia of the oil flow in the piping and the hydraulic machines.

It is therefore possible to have the engine running when the transmission shaft stalls for any reason; as a matter of fact the turbine blades cannot stop the oil and can, in the mean time, receive a force from the flow.

Such transmission shows the following advantages:

- The transmission ratio can be continuously changed as a function of the torque ratio.
- There are no wearing parts.
- There is a high damping capacity for torsional vibrations.
- There is no danger of stalling the engine as a consequence of too high a torque being applied to the gearbox input shaft.

On the other hand we can think about the following disadvantages:

- Low transmission efficiency in all conditions
- Increased transmission complexity

As we have already commented, the converter cannot be easily removed from the driveline and has a significant polar inertia; this makes the application of conventional synchronizers impossible. It is therefore necessary to use powershift gearboxes, where shifting speed can occur in longer time frames, because there is no torque interruption.

The kind of lay-out we have imagined could be too bulky for vehicle application; H. Föttinger had the idea of integrating the pump and turbine in a single compact machine, avoiding connecting pipes.

This result was obtained by using turbine and pump wheels of similar radial dimension. In this case the wheels can face each other, creating the hydraulic circuit without additional elements.

Figure 11.12 shows a diagram of this lay-out; the pump is made with a radial bladed wheel P; the disc supporting the blades shows a cavity where the turbine T can be installed.

The pump is connected to the engine crankshaft. The volume where the oil flows is limited by the rotating discs, the blades and the closure surface C, which can also be avoided.

Turbine and pump are almost identical; the two wheels must be positioned with the same center line and a rotating seal must be provided, to keep the oil inside. The proposed scheme shows no stationary wheel suitable for receiving a reaction torque. Input and output torque must in this case be equal; this particular machine is called an *hydraulic clutch* and acts like a friction clutch.

This machine cannot transmit any torque if there is no speed difference between pump and turbine; in fact, the oil flow rate determines the transmitted

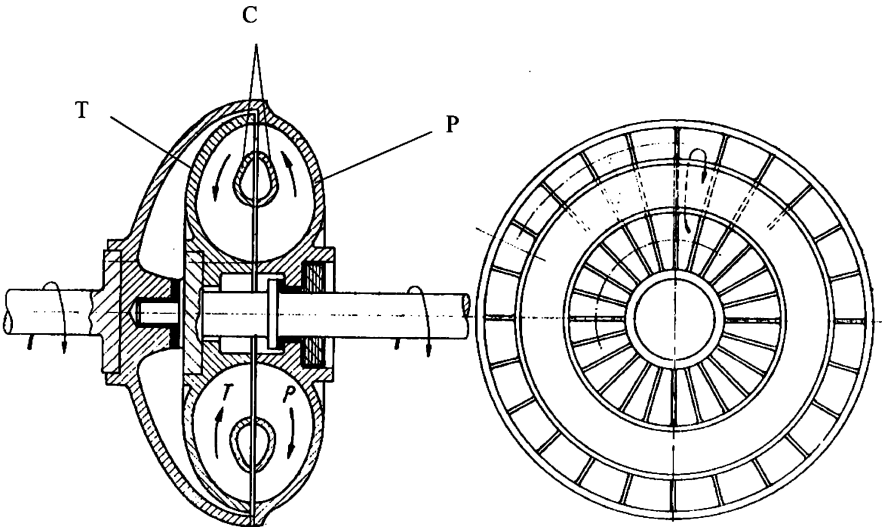


FIGURE 11.12. Hydrodynamic clutch scheme; P is the centrifugal pump and T the turbine. The arrows show the direction of flow when the pump speed is greater than the turbine speed.

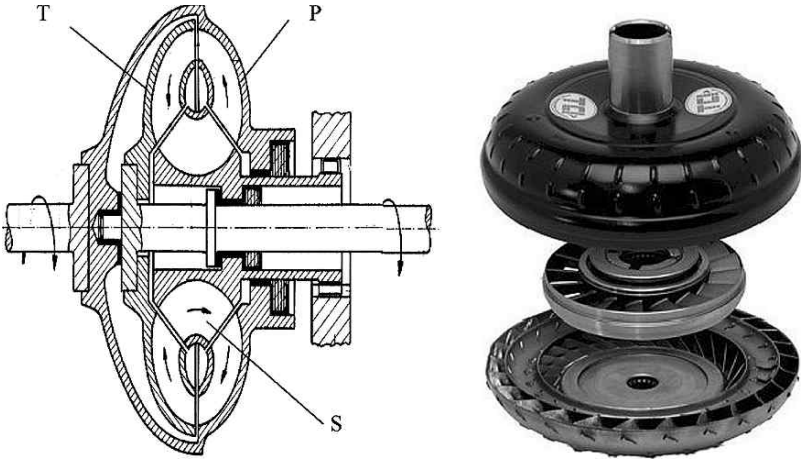


FIGURE 11.13. A torque converter scheme is represented on the left. A picture of the elements of a torque converter is shown on the right.

torque and it becomes zero when speeds are equal. If the wheels had a different geometry, the transmitted torque would be zero when the flow rate stopped.

Figure 11.13 shows instead the scheme of a *torque converter*; unlike the previous one that includes only the pump P and turbine T, here there is a third bladed wheel S, also called the *reactor* wheel, connected to a standing element, such as the gearbox housing.

In this case the torque on the pump and on the turbine might be different, because the reacting element can equal the algebraic difference of torque.

If, as occurs more frequently, the output torque is greater than the input torque, the reaction torque must take the same direction as the input torque. If the input torque should be less than the output torque, the reaction torque should have direction opposite to the input torque.

If the stator is connected to the gearbox housing with a freewheel that can only withstand a torque with the same direction as the input torque, the machine could multiply the input torque or behave like a hydraulic clutch, when the output torque equals or is greater than the input torque.

The picture at the right of the diagram shows a disassembled torque converter made completely out of stamped steel sheet parts.

11.2.1 Hydraulic clutches and torque converters

In order to understand the operation of a rotating bladed wheel, we consider a wheel with a single, almost radial, channel whose walls show angles β_1 and β_2 at the outer intake radius and the inner exhaust radius, as shown in Fig. 11.14.

We call r_1 and r_2 the outer and the inner radii of this wheel.

The wheel is struck by a mass flow rate Q, coming out of a nozzle that is inclined at the angle α_1 with respect to the tangent to the outside diameter

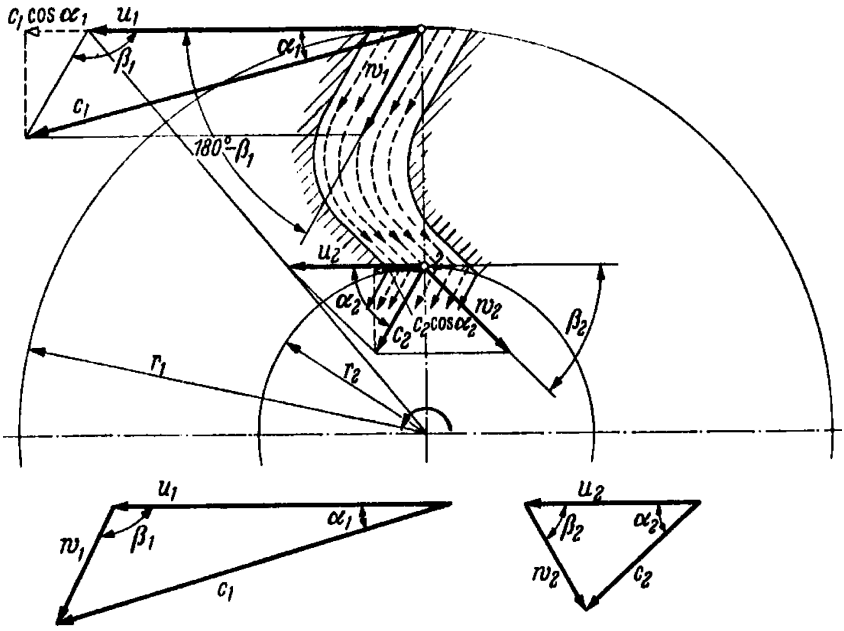


FIGURE 11.14. Speed triangles on a bladed wheel invested by a nozzle with an output speed c_1 inclined at an angle α_1.

of the wheel. We call c_1 the output velocity of the oil from this nozzle; for the sake of simplicity we suppose the speed to be constant at any point of the cross section of the flow.

The torque applied to the bladed wheel will be given by the difference between the mass flow momentum at the intake and the exhaust of the channel on the wheel. The two speed triangles at the input and output cross section of the flow allow us to calculate the absolute speed of the flow when leaving the wheel.

In particular, the input speed c_1 can be considered as the vectorial sum of the oil speed relative to the channel and of the speed of the oil considered as part of the channel; the first speed is w_1, whose direction is parallel to the walls of the channel at the intake cross section of the channel. The second speed is the peripheral speed of the wheel at the radius r_1, which we call u_1.

The absolute speed at the exhaust cross section c_2 can be calculated starting with the speeds u_2 and w_2 at the exit of the channel; notice that the first is defined by the radius r_2, while the second is derived from the continuity of the mass flow in the channel: its direction is again tangent to the walls of the channel and its module will be equal to the module of w_1 multiplied by the ratio of the areas of the two cross sections.

The momentum at the intake of the wheel is:

$$Qc_1 \cos \alpha_1 r_1 \ ,$$

while the momentum at the exit of the wheel is:

$$Qc_2 \cos \alpha_2 r_2 \ .$$

The torque on the wheel will be therefore:

$$M = Q(c_1 \cos \alpha_1 r_1 - c_2 \cos \alpha_2 r_2) \ . \tag{11.2}$$

It is sufficient to know the intake and exhaust angles of the blades and the related radii to determine the torque on the wheel with the approximations that we have introduced. The case of the hydrodynamic clutch and of the torque converter is complicated by the fact that the absolute speed of the intake is unknown, because it is bound to the speed of the preceding wheel.

Let us consider, for instance, the hydrodynamic clutch that is introduced in Fig. 11.15. P is again the pump, while T is the turbine; the blades are perfectly radial.

We assume that the oil is incompressible and the current line is perpendicular to the separation plane between the two wheels and, therefore, parallel to the axis of the wheels. The fluid moves according to organized vortices with their axis perpendicular to the axis of the wheels.

If we consider the separation plane between the two wheels, we could draw on it a circumference (the dots and lines in the figure) separating two constant flows of opposite direction.

The cross section between the blades that are inside and outside of this circumference will have the same area, because the oil is incompressible.

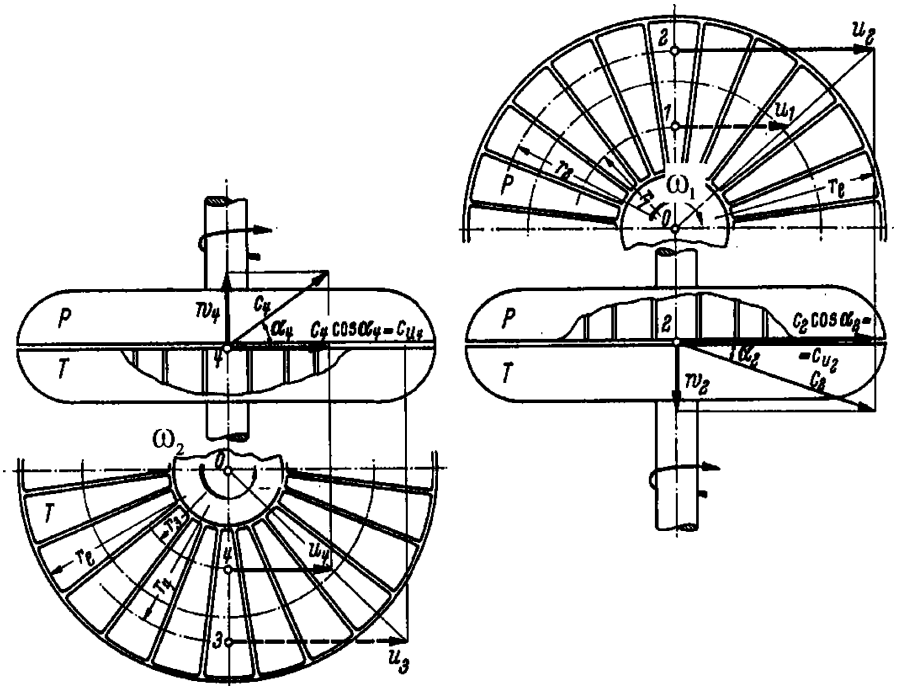


FIGURE 11.15. Speed triangles in a hydraulic clutch.

If the speed is constant on the cross section we can calculate the speed on the centres 1 and 2 for the pump and 3 and 4 for the turbine; the calculation can be repeated with the same results for all the other blades.

In a generic operation point the speeds Ω_2 and Ω_1 of the turbine and the pump will not be equal; more likely is that, because the pump is the moving element, its speed will be greater than that of the turbine. On the facing points 1 and 4 at the intake of the pump, the speed c_1 will be determined by the two components:

$$w_1 \text{ and } \quad u_1 = \Omega_1 r_1.$$

Be careful not to confuse the subscripts of the angular speeds bound to turbine and pump, with those of oil speeds bound to the passage sections.

In the same way the speed triangle at the facing turbine section can be built; it is shown in the figure and takes into account the fact that the two components w_1 and w_4 are equal because the cross sections are equal.

The vectorial difference:

$$\mathbf{c}_1 - \mathbf{c}_4 = \mathbf{u}_1 - \mathbf{u}_4 \tag{11.3}$$

represents a part of the mass flow momentum of the pump that cannot be utilized by the turbine, because of the difference of rotational speed of the two wheels.

In the same way we can proceed in sections 2 and 3, if we again remember that the components w are equal to those of the sections 1 and 4.

At the intake of the pump, at section 1, there is an energy loss due to the fact that the fluid is given by the turbine at a peripheral speed lower than that of the pump; the latter must provide for a possible sudden acceleration. This type of loss is called *impact loss*. In the same way at the intake of the turbine, at section 3, there is a sudden deceleration of the flow, produced by the same cause; if we consider the impact losses the speed triangles of the two wheels will be the same at points 1, 4 and 3, 2.

The torque on the wheels will be:

$$M_1 = M_2 = Q\left(u_2 r_2 - u_4 r_1\right) = Q u_2 r_2 \left(1 - \frac{u_4 r_1}{u_2 r_2}\right) = Q u_2 r_2 \left(1 - \frac{\omega_2 r_1 r_1}{\omega_1 r_2 r_2}\right) ,$$

or:

$$M_1 = M_2 = Q u_2 r_2 [1 - \nu(\frac{r_1}{r_2})^2]. \tag{11.4}$$

If we calculate the power from this last equation, we again obtain the equation 11.1, where $\mu = 1$:

$$\eta = \frac{P_2}{P_1} = \nu.$$

The above formulas do not take friction losses of blades in the oil and windage losses on wheels into account. These last are negligible until the transmitted torque is high. Friction losses in the oil, called *hydraulic losses*, can be imagined as a function of the square of the flow speed w, according to the formula:

$$P_a = kQw^2, \tag{11.5}$$

where k is a constant depending upon the nature of the surface of the blades.

In cases where the wheels have the same rotation speed, the pressure value they generate on the fluid at the outside boundary between the wheels are equal and opposite. In this case there will be no oil circulation and the flow rate must go to zero; as a consequence the processed torque and power are also zero.

The case of a torque converter could be treated in the same way.

In Fig. 11.16 the diagram of a torque converter is drawn; on the schematic cross section at the right of the figure the dotted lined represents the center line of the channel between adjacent blades, joining the centres of the different cross sections.

These centers are determined by the discs and blades of the pump P, the stator S and the turbine T and by the internal walls that are now present; the position of these walls is such as to obtain an almost constant circulation speed w.

We imagine drawing traces of blades on the rotation surface generated by this line around the converter axis, and developing this surface on a plane. Thus we can arrive to the scheme in the centre of the figure.

On the right side of the same figure there are speed triangles that can be built just as they were built for the hydrodynamic clutch.

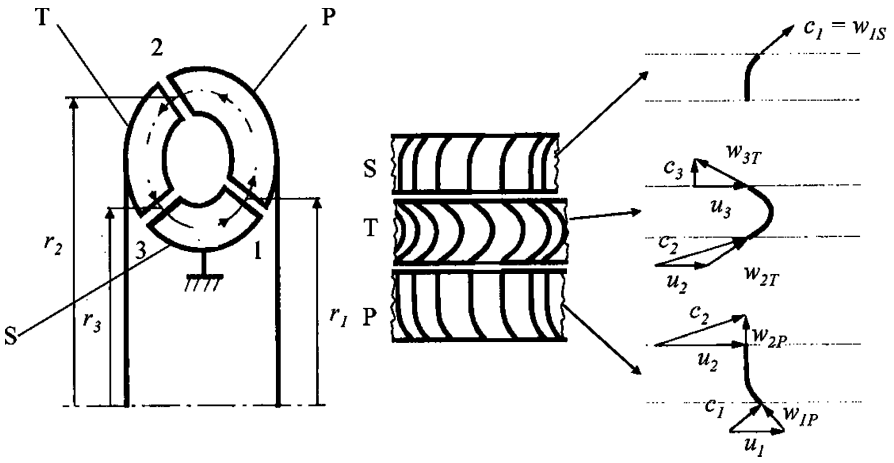


FIGURE 11.16. Speed triangles in a torque converter.

Sections 1, 2, 3 are relative to the transition points between stator and pump, pump and turbine, turbine and stator. To the speed w is given a second subscript, according to the converter element to which each refers.

At the transition point between one element and the next, speeds c are conserved, speeds u are calculated from the radius r of the centre of the section they refer to, and speeds w are tangent to the blade walls.

Because the module of w is determined by the continuity equation of the mass flow, only in one condition shown in the figure will there be no impact losses. This condition is called the *design point*. In these conditions the efficiency will be maximum and determined only by hydraulic losses. At conditions different from the design point, the geometry of the blades will cause also impact losses.

It is usual for torque converters and hydrodynamic clutches to use the principles of similarity.

The reference geometric dimension is a diameter, usually the outer diameter of the largest wheel D; if ρ is the mass density of the fluid we can define a performance coefficient λ defined by the formula:

$$M_1 = \lambda \rho \omega_1^2 D^5. \tag{11.6}$$

Two machines are *similar* if their linear dimensions are in scale and blade angles are equal; if two machines are similar, the torque absorbed by the pump will be defined by the same *performance coefficient* λ.

The non-dimensional entities such as η and μ in similar machines will also be equal at the same value of ν.

11.2.2 Characteristic curves

Figure 11.17 represents in a qualitative way the *characteristic curves* of a hydrodynamic clutch; these characteristics include the performance coefficient λ

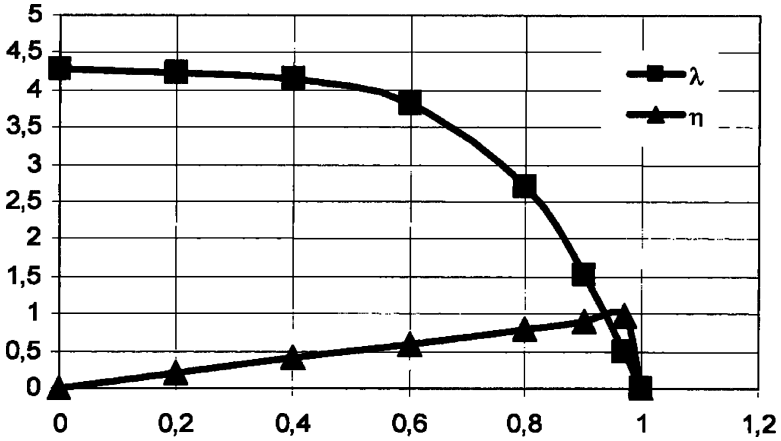


FIGURE 11.17. Characteristic curves of a hydraulic clutch representing λ and η as a function of ν. λ is multiplied by 10,000.

and the efficiency η. It should be observed that the performance coefficient is multiplied by 10,000.

λ decreases with the speed transmission ratio, because the processed torque is dependent upon the flow, as determined by the speed difference of the wheels; when the wheels are synchronous, $\nu = 1$ and the torque is zero. The efficiency should always be equal to the speed transmission ratio except at $\nu = 1$. At this point it also goes to zero.

This fact is justified if we take windage losses and rotating seal friction into account; because the processed power at synchronous speed is zero, even small losses will cause the efficiency to go to zero, too.

In a hydraulic clutch, the torque transmission ratio is equal to 1, by definition.

Figure 11.18 represents instead the characteristic curves of a torque converter; characteristics include in this case λ, η and μ, the torque transmission ratio. It should be observed that, as before, λ is multiplied by 10,000.

Because there is a reaction element, the torque transmission ratio is not constant; the torque absorbed by the pump does not change significantly as the speed transmission ratio increases, because of the reactor wheel.

The torque transmission ratio starts from a value of about 3–4 when the speed transmission ratio is zero, when the output shaft is stalling, and goes to zero when the input and output shaft are synchronous; as a consequence, the efficiency that is the product of the two transmission ratios will be zero as $\nu = 0$ and $\nu = 1$, where $\mu = 0$.

In the same figure some typical operating conditions are shown:

- S, standing for *stall*, is the condition where the turbine is still and the pump is rotating.

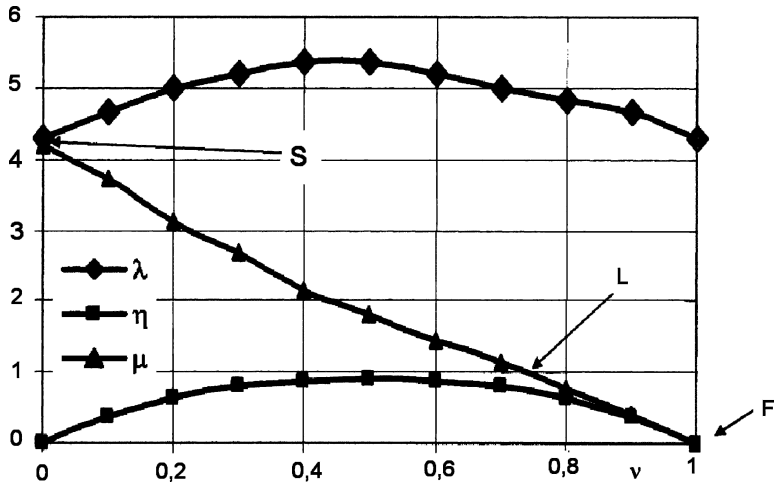


FIGURE 11.18. Characteristic curves of a torque converter representing λ, η and ν as a function of ν. λ is multiplied by 10,000.

- L, standing for *lock-up*, is the condition where the torque on the pump and the torque on the turbine are equal, so that the torque transmission ratio is one.

- F, standing for *free flow* point, is the condition where torque on the turbine is zero.

All curves can be justified by the construction of speed triangles and the principles of similarity.

The Trilok torque converter, named after the research consortium that was in charge of its development, provides for a reactor wheel mounted on a free wheel that can react only to those torques having the same direction as the input engine torque. With this provision, at the lock-up condition the reactor wheel is free; in fact, from this point on, the turbine should receive a lesser torque than the pump and, as a consequence, the reactor wheel should experience a torque opposite to the engine torque.

If the reactor is free wheeling, the torque converter will behave as a hydrodynamic clutch; therefore the efficiency, as shown in Fig. 11.19, will become linear at point L, with a steep decrease to zero at point F. From point L on, the torque transmission ratio will also remain constant, instead of going to zero.

As a first approximation, Trilok torque converters show, in any operating condition, the best efficiency between the hydrodynamic clutch and the torque converter.

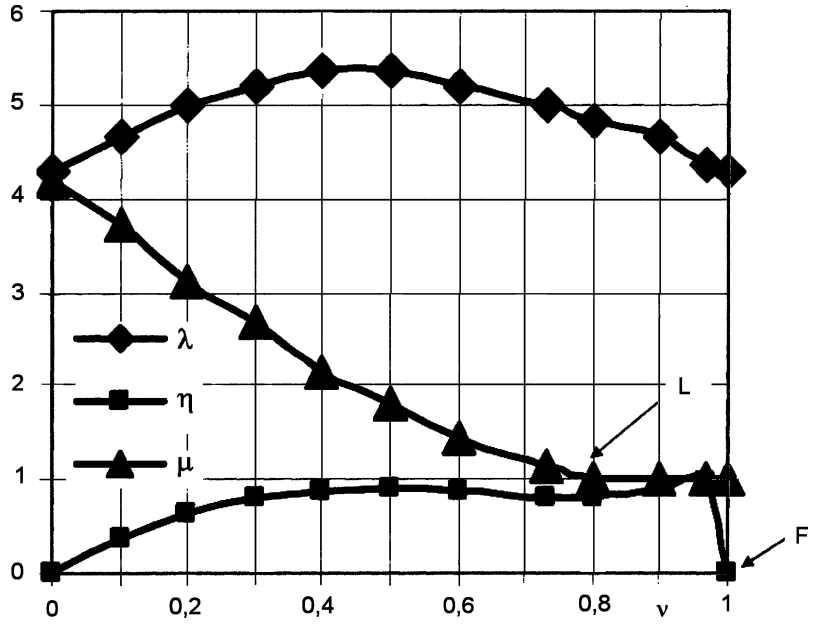


FIGURE 11.19. Characteristic curves of a Trilok torque converter representing λ, μ and η as a function of ν. λ is multiplied by 10,000.

11.2.3 Torque converter performance on a vehicle

Matching a torque converter to an engine and a vehicle implies the performance calculation of a system including engine, torque converter and vehicle; this calculation can be made taking into consideration a diagram where the engine torque curve and pump torque curve are represented. Because pump and engine rotate together, the engine speed also represents the pump speed.

Once the torque converter diameter is chosen, λ has been defined; the torque absorbed by the pump is a parabola.

We can start drawing the diagrams represented in Fig. 11.20. These represent the torque absorbed by the pump at different values of the speed transmission ratio, the engine output torque at wide open throttle M_{max} and a torque curve M_{reg} at partially open throttle.

As a value for the speed transmission ratio is assumed, for the equilibrium of the pump and engine system, the engine speed will be forced at the value found at the intersection point of the engine torque and pump torque diagrams.

In order to change the matching point, the engine torque must be changed.

Let us remember that the fact that the engine can deliver the maximum torque at stall is bound to the choice of pump diameter.

From torque converter characteristic curves, the gearbox input torque (torque converter output torque) M_2 can be calculated as the engine torque M_1 is known.

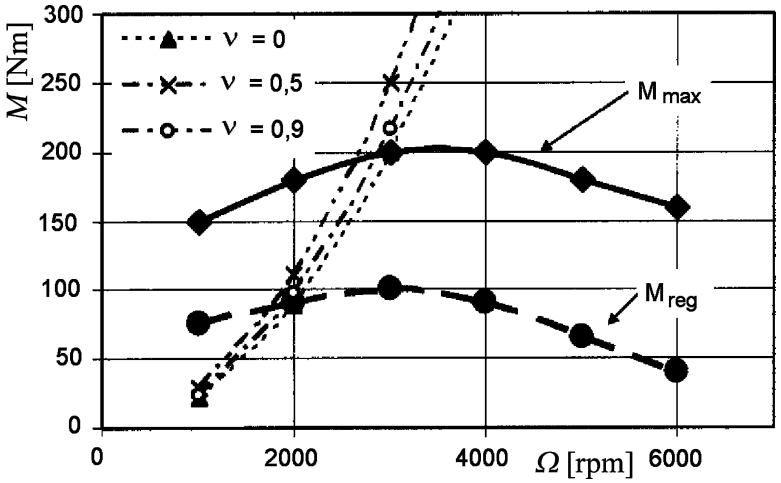


FIGURE 11.20. Matching torque curves of the engine and of a torque converter, at different values of the speed transmission ratio ν. Notice that the operating speed is determined by the engine torque and the speed transmission ratio only.

The gearbox input speed Ω_2 will be given by the product of the engine speed Ω_1 by the speed transmission ratio ν; the gearbox input torque M_2 will be given instead by the product of the engine torque M_1 by the torque transmission ratio μ.

The calculation can be performed starting from the diagrams of Figs. 11.19 and 11.20, for the wide open throttle engine torque; the procedure can be repeated for any values of regulated engine torque.

The characteristic of the engine transmission system is shown in Fig. 11.21; observe how the very small section of the engine torque curve is expanded to the gearbox input torque curve.

Two other cases are shown in the same diagram:

- The case of an ideal continuously variable transmission with mechanical efficiency equal to one and
- The case of a friction clutch with a safety margin of 15%

The converter performance lies between the two reference curves, closer to the ideal continuously variable transmission.

The traction curve of the vehicle T can now be drawn, as a function of the vehicle speed V, multiplying M_2 and Ω_2 by the different values of the gearbox transmission ratios, utilizing the well known formulae:

$$V = \Omega_2 R_r / \tau_i \ , \quad \text{with} \quad i = 1, n \ ,$$

$$T = M_2 \tau_i \eta_t / R_r \ , \quad \text{with} \ i = 1, n \ ,$$

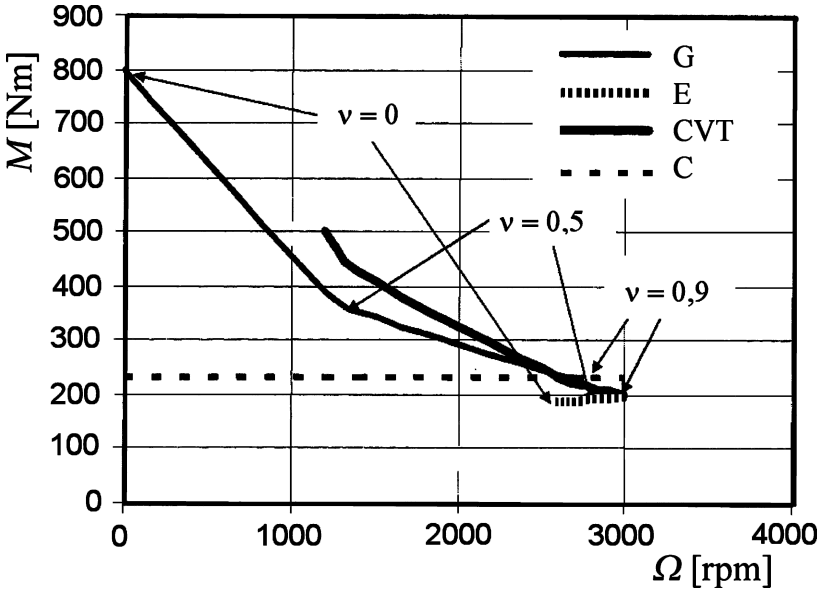


FIGURE 11.21. Calculation of the gearbox input torque M_2 (G) starting from the engine torque M_1 (E). The values of the speed transmission ratio of the different curves are shown. The values of M_2 are also reported for an ideal continuously variable transmission (CVT) and for a slipping friction clutch (C).

where τ_i is the total transmission ratio of each speed of a gearbox with n speeds, R_r is the wheel rolling radius, η_t is the transmission mechanical efficiency, not including the torque converter.

The time integration of the available power, according to the methods that will be explained in the second volume, can be made in order to calculate the performance of the vehicle with a torque converter.

By the repetition of this procedure for a number of values of partially open throttle torque, the fuel consumption map of the engine and torque converter system can also be calculated; for each value of the engine torque, the associated value of the fuel consumption can be divided by the corresponding value of the torque converter efficiency.

The use of this new fuel consumption map can lead us to calculate the consumption of the vehicle when the torque converter is working.

Sizing the torque converter in such a way that the engine converter system can stall at engine maximum torque does not always emerge as the best design praxis. It is convenient to look as well for different sizing criteria, with torque converters with higher and lower diameters.

Higher diameters will force the engine to work at lower speeds, lower diameters at higher speeds.

The consequent change of the stall torque will also cause the traction curve to change at wide and partially open throttle. The start-up transients will show

a different duration and also different consumption. A trade-off between performance and consumption can be studied.

We can say that the contribution of a torque converter can be compared with that of an added range changer gearbox with variable efficiency and transmission ratio.

The poor efficiency of the torque converter has suggested using a *lock-up clutch*. This is done with a multi-disc oil clutch whose closure is controlled by the gearbox control system; when this clutch is closed the torque converter behaves as a rigid joint.

Closure of the clutch occurs when the pump and turbine speeds are close to each other; in these conditions the torque converter contributes little to the performance and compromises efficiency.

It must not be forgotten that by closing the lock-up clutch, the mechanical damping function of the torque converter is also lost; a spring damper is therefore necessary. Sometimes it can be convenient to allow a controlled slip on the clutch, in order to avoid torsional vibrations.

The closure of the lock-up clutch must therefore be gradual at all times. In some automatic transmissions of previous generations this function is inhibited in the lower gears; in modern transmissions the clutch is also modulated in order to avoid vibration without jeopardizing the advantage for fuel consumption.

12
SYNCHRONIZERS

12.1 DESCRIPTION

The function of a synchronizer is to enable meshing gears to be changed, on a moving vehicle without negative consequences for gears mechanical integrity and interior noise.

During synchronization the friction clutch must be disengaged.

12.1.1 Simple synchronizer

The most widely diffused synchronizer is the Borg Warner, named after the manufacturer that developed it. Figure 12.1 shows a cross section of this synchronizer, in its single cone version; there are, as usual, two synchronizers, for managing two neighboring gears on the same shaft.

The gears are idle and are mounted on their shaft on needle bearings; they could be mounted either on the input or output shaft and are constantly meshed with those fixed on the other shaft.

The two idle gears show on their adjacent side the synchronizer *hub* 3, fixed to the relative gear. The joint between the hub and the wheel can be made in different ways: By laser welding, as in the figure, or by a spline shaft. Teeth can also be directly cut or stamped on the side of the gear.

In this last case, there must be a sufficient difference in radial dimensions between the gearing teeth of the wheel and the selector teeth of the synchronizer hub.

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

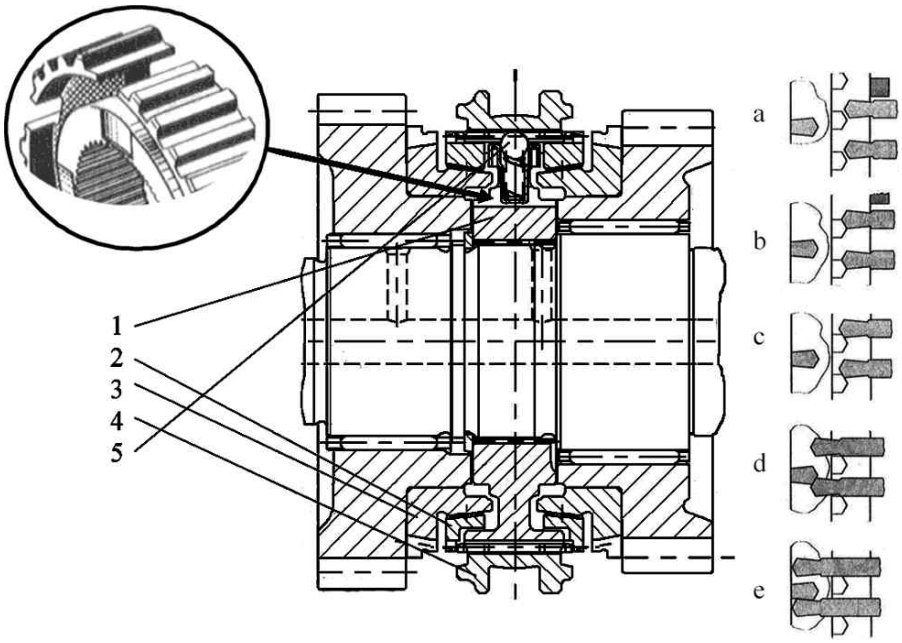


FIGURE 12.1. Cross section of a single cone Borg Warner synchronizer (FIAT). In the circle at the upper part of the picture there is a detailed view of the synchronizer body 1. On the right the positions of the teeth of the hub 3, of the ring 2 and the sleeve 4 are sketched, in the different phases of the synchronization process.

The different fabrication technologies of the joint between the hub and the wheel have a direct consequence for the length of the synchronizer assembly and the gearbox itself.

The hub 3 shows a crown of teeth on its outer circumference; the *synchronizer ring* is free to move axially and matches with the hub through a tapered friction surface. It can also match with the selector teeth of the sleeve 4, in such a way as to block the rotation but leave the axial motion free.

The synchronizer ring 2, between the hub and the sleeve, acts as a friction clutch that can synchronize the elements before the engagement of the teeth; how this function can be achieved correctly will be discussed in a following paragraph.

12.1.2 Multiple synchronizers

Single cone synchronizers were widely used until few years ago; in recent times the target of shortening shift time without increasing the manual force applied by the driver has taken priority. Multiple cone synchronizers have therefore come into use, particularly on low gears where the synchronization work is more demanding.

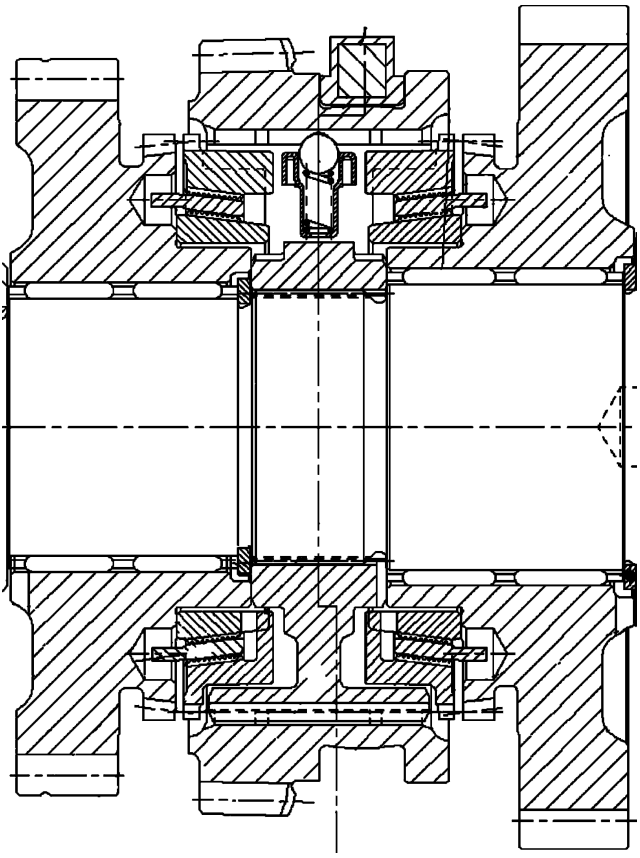


FIGURE 12.2. Cross section of a double cone synchronizer (FIAT).

Figure 12.2 shows a double cone synchronizer; this arrangement can be appropriate to the mid-gears of a large gearbox or the lowest gears of a small one.

Let us look at the different arrangement of the tapered rings of the synchronizer hub. The hub is reduced to the teeth crown only but there is a tapered friction ring connected to the hub in such a way as to rotate with it, allowing some freedom to move in other directions.

Friction surfaces are machined on the inside and outside faces of this ring.

The matching tapered surfaces are made with two different elements fixed to the selector sleeve in a particular way. The outside surface has the same shape as a single cone synchronizer, as we will explain in detail later; the inside surface is matched to the outside through a spline shaft that does not rotate but leaves a certain freedom along the radial and axial directions.

This freedom is necessary to compensate for the wear and alignment errors of the friction surfaces.

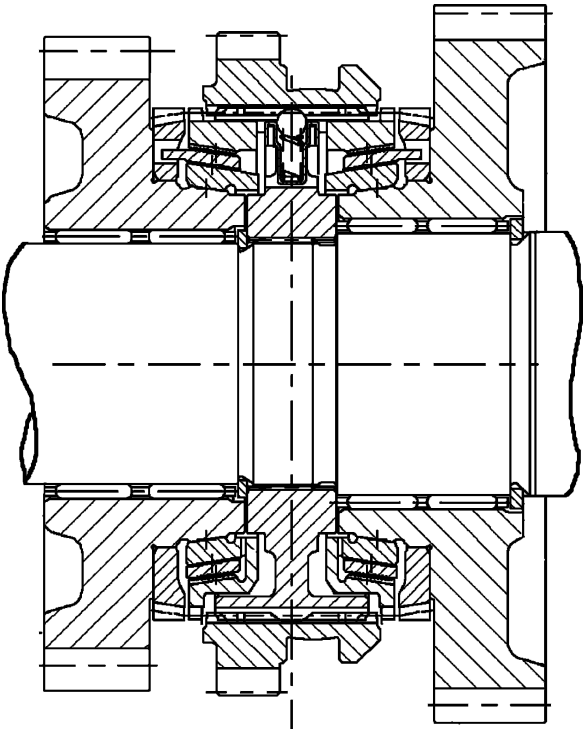


FIGURE 12.3. Cross section of a triple cone synchronizer (FIAT).

In the following Fig. 12.3 two triple cone synchronizers are shown; this arrangement is appropriate to first and second gears.

In this case active friction surfaces of the hub are three; the innermost is made like that of a single cone synchronizer, while the remaining two are similar to those of a double cone synchronizer.

Active sleeve surfaces are respectively arranged as a single cone synchronizer (on the outside) and mounted on a spline with some axial and radial clearance (on the inside).

The advantage of these arrangements is the multiplication of the synchronization torque made by the number of active friction surfaces present on the synchronizer; this allows the containment of the diameter of the ring, as compared with a single cone synchronizer of the same performance.

It is possible to synchronize gearshifts without using synchronizers. Rotating parts should be accelerated during downshifts through the engine, after having engaged the clutch, when the gearbox is in neutral. It should also be possible to slow down rotating parts by means of a brake installed in the gearbox. This solution could prove of interest for the automation of manual transmissions by electro-hydraulic electronic control systems.

12.1.3 The gearshift process

Let us examine the implications of synchronizing a gearbox during a speed shift; we will refer to a simple two speed single stage gearbox, shown in Fig. 12.4; power flows through the input shaft e to the output shaft u, through two alternate gear couples.

To shift speed, the sleeve must alternately engage with one of the wheels, in positions 1 and 2.

When couple 1 is engaged, if we call z_{1e} and z_{1u} teeth numbers on the two meshing wheels and Ω_e and Ω_u shaft rotation speeds, we will have:

$$\Omega_u = \frac{\Omega_e z_{1e}}{z_{1u}} \ , \tag{12.1}$$

while the speed of the idle gear of couple 2 will be:

$$\Omega_{e,2} = \Omega_e (\frac{z_{1e}}{z_{1u}})(\frac{z_{2u}}{z_{2e}}) \ . \tag{12.2}$$

If we shift the sleeve to position 2 immediately, also after clutch disengagement, a number of shocks will occur between the dog teeth of the two wheels; noise and structural damages could result.

Synchronizing parts means bringing the speed difference of moving parts to zero, before positive engagement. Because it is obviously impossible to change the speed of the u shaft, or the vehicle, without the help of the engine (the

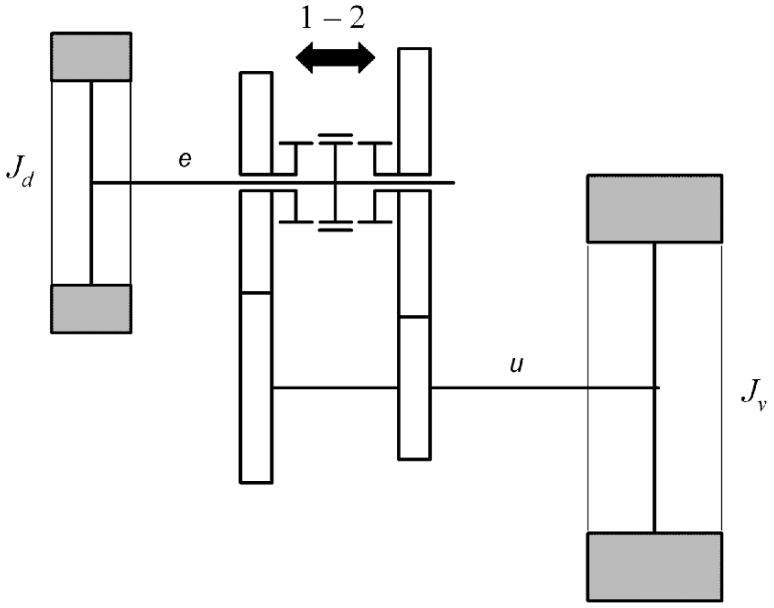


FIGURE 12.4. Diagram of the masses to be synchronized during the speed shift of a two speed single stage gearbox.

clutch is disengaged), synchronizing also means changing the rotation speed of parts bound to shaft e in such a way as to eliminate the speed difference of the synchronizer body and the tooth wheel, before dog clutch engagement.

In our example, shifting from first to second speed implies reducing the speed of e, while during down shifts the same speed must be increased.

This operation requires a certain time delay, during which some energy must be added to or subtracted from the system. In the mean time, taking into account the total driving resistance, the vehicle speed and the rotation speed of the shaft u will decrease. On a downhill road the opposite may occur.

Rotating masses whose speed must be changed are represented by a flywheel of inertia J_d; to its value contribute: Driven clutch plate, gearbox input shaft with its fixed wheels (the constant gear wheels and the complete countershaft with its fixed wheels, in case of double stage gearboxes).

During a shift from second to first, some energy must be introduced into the system, while in the case of the shift from first to second, some energy must be subtracted; synchronizers perform this function through friction rings between the two masses to be brought to the same speed.

The necessary energy for this manoeuvre is subtracted or added to the apparent rotating mass of the vehicle, represented by a flywheel of inertia J_v rotating at the speed Ω_u. A vehicle deceleration or acceleration will therefore perform the work of synchronization. Because the inertia J_v is much greater then J_d the vehicle speed variation will be negligible.

We can therefore summarize the functions performed by the synchronizer as follows:

- Adapting the speed of parts to be synchronized by an energy transfer from one to the other. The measure of this function is the synchronization time; the latter is a significant part of shifting time during which the vehicle is without traction.

- Positively joining synchronized parts so as to transmit the necessary torque; the joint must be stable in time with no danger of gear self-disengagement.

In addition to the above, the following additional functions must be assured:

- Measuring the speed difference of rotating parts, to identify the suitable moment to engage the speed

- Enabling positive engagement only when speeds are equal

Refer again to Fig. 12.1 and examine additional details.

A magnification of the synchronizer body is shown at the upper left of the figure. The body 5 shows a spline surface with six interruptions at 60 deg ; these interruptions are replicated on the synchronizer ring 2.

Ring interruptions match, with a defined angular clearance, with three large teeth on the inside diameter of the sleeve.

The remaining three interruptions at 120 deg match with three push elements 5 that are free to slide inside the three said elements, which are the sleeve, the body and the ring; the push elements show a radial hole in which a plunger with spherical head is fitted and is pushed by a coil spring against a seat on the inside of the sleeve.

Assume that the gearbox is in neutral, with both engine and vehicle in motion. The left wheel must be engaged.

We can identify five different phases of the shifting manoeuvre of the sleeve, actuated by the shift stick; for each of these phases a schematic cross section is drawn, showing a teeth view of sleeve, ring and hub developed on a plane.

1. The sleeve is pushed to the left against the reaction force of coil springs that try to keep the push elements 5 in neutral position. Elements 5 and large teeth push the ring 2 and bring into contact the two matching tapered friction surfaces of ring and hub. Because the two parts have different speeds, a friction torque will arise against ring and hub relative speed. The circumferential clearance between the large teeth and ring will allow a small rotation of the latter; the clearance is designed in such a way that the ring teeth stop the axial motion of the sleeve. See the detail a at the upper right. This phase is called *pre-synchronization* because it does not imply a real energy exchange between moving parts.

2. Friction torque between ring and hub transfers energy from the heaviest part of the system (the vehicle) to the lightest, in order to slow or to accelerate rotating parts fixed to the wheel to be shifted. Friction torque *measures* the relative angular speed because it will go to zero with relative speed. The same torque allows sleeve motion to stop until synchronization is complete. This phase is called *synchronization*.

3. Before describing the next phase, let us consider the shape of dog teeth on hub and sleeve. These show a tapered end and a counter tapered surface. When ring and hub are synchronous the friction torque is zero and the tapered ends of the dog teeth rotate the hub of the angular thickness of the half tooth, under the pressure exerted by the shift stick.

4. The sleeve is now free to move across the ring and to match the hub; again the tapered end of the teeth will rotate the hub if necessary. In this phase the push element of plungers are completely retracted and no longer withstand sleeve motion.

5. Now a *positive engagement* is made; at the end of this phase the driver will stop pushing shift stick. The counter tapered shape of the dog teeth will retain the sleeve in the engaged position under the action of the engine torque. The same shape will exert a certain reaction on the shift stick, when the driver begins to disengage the gear for the next speed shift.

12.2 DESIGN CRITERIA

Synchronization torque determines the length of the shifting operation, which contributes to the time needed for vehicle acceleration.

This torque is proportional to the force applied to the shift stick; the latter must respect certain design conditions.

The force on the shift stick knob must be kept below 80–120 N on cars, while 180–250 N can be allowed on industrial vehicles; synchronization time must be lower than 0.15–0.25 s on cars, while on industrial vehicles it must be lower than 0.25–0.4 s.

Transmission ratios between the shift stick knob and sleeve must be kept below 7–12:1, to avoid excessively long shifting strokes.

When calculating the force available at the synchronizer, beginning with the force of the lever, mechanical efficiency of shifting mechanisms must be taken into account; refer to our discussion of the clutch mechanisms.

Figure 12.5 shows synchronization times versus shifting force applied to the stick knob, for several manual single stage gearboxes, for shifts from fourth to fifth and vice versa; the values for up and down shifts are not the same, because the synchronization works are different. The more effective the design, the shorter the synchronization time at the same shifting force.

The force applied to the knob is not the parameter for judgment. The shape of the shift force diagram versus the lever stroke is very important, too; Fig. 12.6 shows diagrams of two different gearboxes.

Satisfactory behavior is shown at the upper part of the figure by a force diagram that generally increases with the stroke, featuring a first peak m_1 associated with the synchronization phase and a second peak m_2 at the end of the engaging phase.

With this kind of mechanical response, drivers can perceive the beginning of the synchronization phase by the force increase; the end of this phase is marked by a sudden decrease of the reaction force. Drivers understand when the shift stick must be kept in place and when the shifting manoeuvre can be completed quickly.

A second positive feature of the first diagram is the statistic consistency between the results of different manoeuvres; this consistency allows the driver to learn proper gearbox use quickly.

Unacceptable behavior is shown in the lower part of the figure, where the phenomenon of the *double shock* is also visible.

The double shock is measured by a third peak m_3; this occurs when the ring must turn at the end of the synchronization phase, aligning its teeth with those of the sleeve.

This phenomenon may be due to an excessive friction coefficient between the tapered ends of the teeth of ring and sleeve or of sleeve and hub. An unsuitable opening angle of the tapered end or a different opening angle on different teeth can be the source of this inconvenience. A difference of opening angle between teeth can also cause different behavior in different manoeuvres.

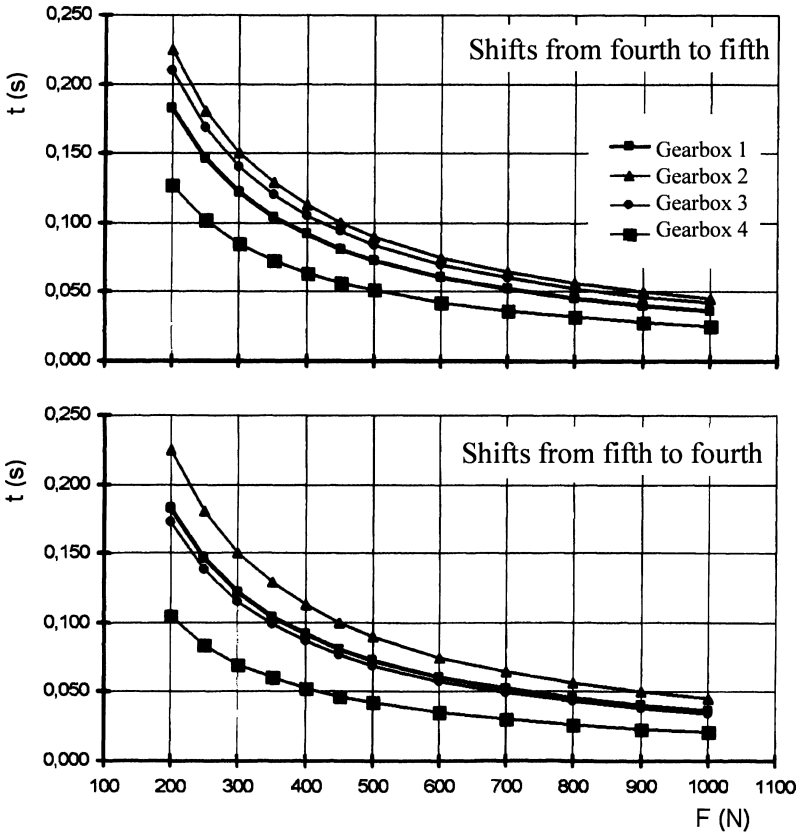


FIGURE 12.5. Diagrams representing synchronization time t (on the vertical axis) versus knob shifting force F (on the horizontal axis), for different single stage gearboxes and for shifts from fourth to fifth and vice versa.

The double shock can lead drivers to manoeuvres characterized by an incomplete shift. The poor statistical consistency of the shifting force is particularly disagreeable because it makes the behavior of the gearbox unpredictable.

Reexamine what was said about gearbox shifting quality at the end of chapter on shifting mechanisms.

12.2.1 Geometric criteria

Ring and hub are represented with a simplified drawing in Fig. 12.7. Mean diameter d and opening angle α of the tapered surface are shown.

Ring and hub must be in contact only on their conical surfaces; it is necessary to provide an axial clearance between parts to compensate for the allowed surface wear during the design life time of the gearbox.

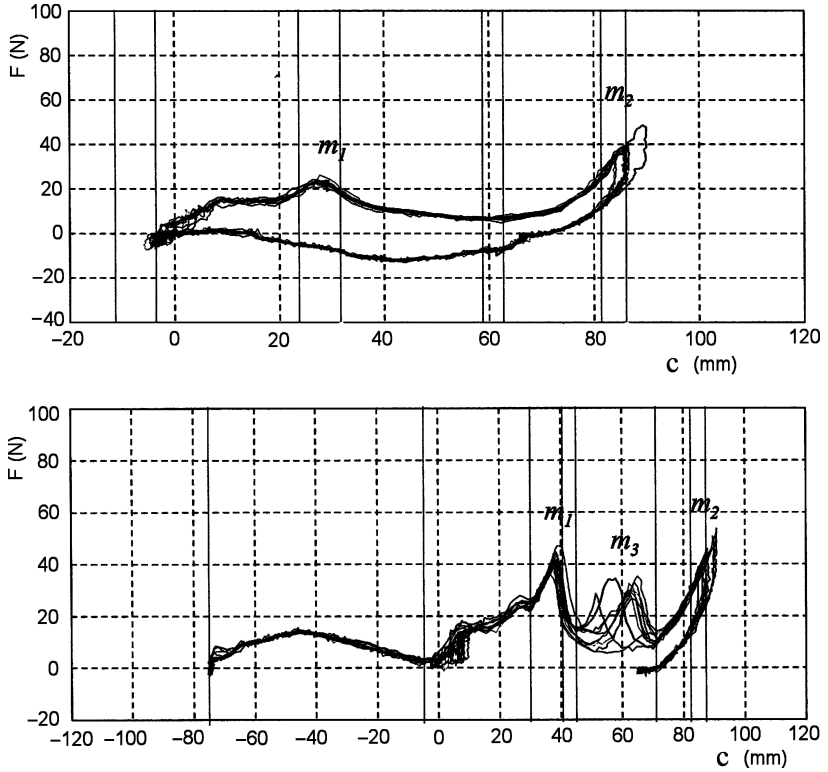


FIGURE 12.6. Diagrams representing shift stick knob force F (vertical axis) versus lever stroke c (horizontal axis) on a single stage gearbox with acceptable and unacceptable behavior (acceptable at the top, unacceptable at the bottom): In the second diagram the double shock phenomenon is announced by the third peak m_3.

Any force F applied to the sleeve originates a normal pressure p on the contacting surfaces; according to wear theory, this pressure is constant, and if we call b the conical surface width, it will be:

$$pb\pi d \sin\alpha = F \ . \tag{12.3}$$

If we designate as f the friction coefficient between surfaces and as M_s the synchronization torque, we can write:

$$M_s = 1/2 \ pb\pi d^2 f = \frac{F df}{2 \sin\alpha} \ . \tag{12.4}$$

Therefore synchronization torque will increase, at the same shift stick knob effort, if the friction coefficient increases. The opening angle will decrease.

But the opening angle cannot be reduced indefinitely or irreversible sticking of the cones will occur.

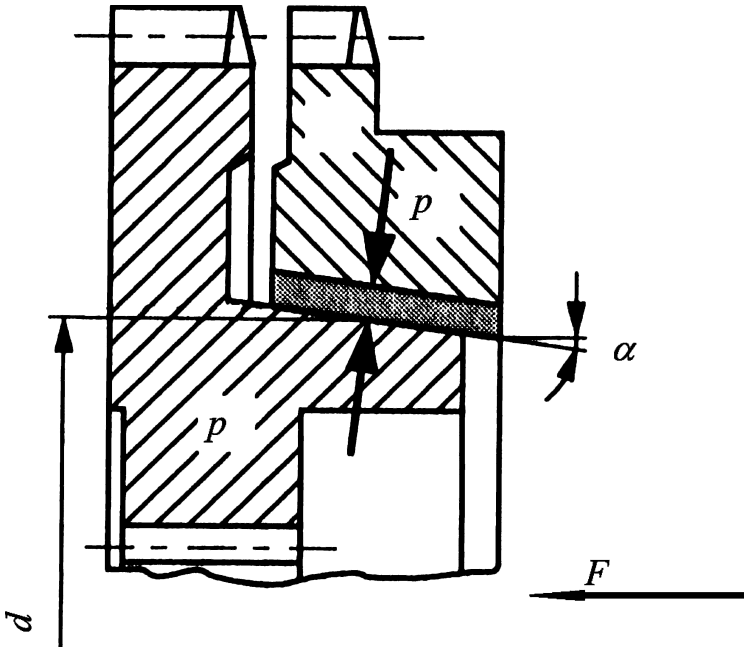


FIGURE 12.7. Scheme of the force F acting on the synchronizer ring, and of the pressure p between conical surfaces.

This happens when the self-unlocking force is lower than the friction force along the unlocking direction; the self-unlocking force is performed by the normal pressure resultant along the axial direction. Boundary sticking occurs when:

$$p \tan \alpha = pf \ , \tag{12.5}$$

therefore rings can unlock spontaneously if:

$$\tan \ \alpha > f \ .$$

The only effective way to increase available synchronization torque is to increase the ring diameter or the number of active surfaces; this last possibility is exploited on multiple cone synchronizers, as shown in this chapter.

Figure 12.8 represents a schematic drawing of a single cone synchronizer, which shows the different contribution to the total axial dimension of the assembly a; contributions to be considered are the following:

- Accepted wear on working friction surfaces; a reference value can range between 0.1 and 0.15 mm; for a conicity of 1:10 the consequent axial clearance must be greater than 1–1.5 mm (Δs).
- Gear torque that determines the width of hub dog teeth.
- Synchronization torque that determines ring dog teeth width.

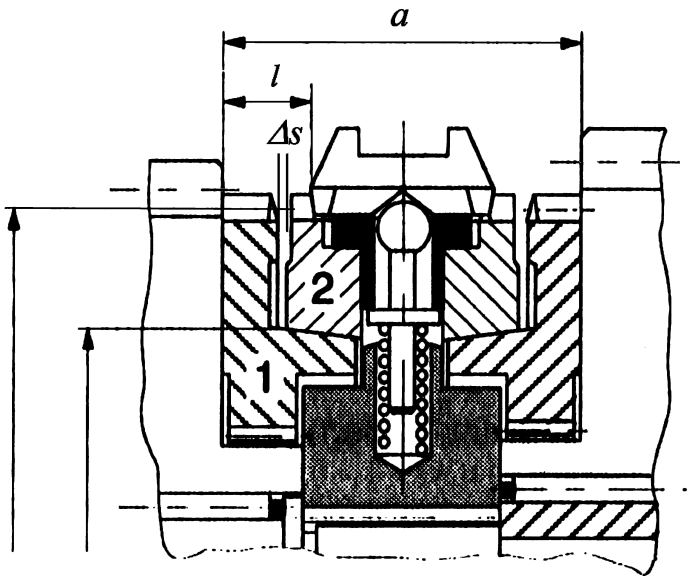


FIGURE 12.8. The axial dimension a of a double single cone synchronizer is determined by the design wear projected in the axial direction and by the dog teeth width of the hub, ring and body.

- The geometrical condition of sleeve engagement and disengagement; sleeve width must be at least double the dimension l coming out from the ring and hub teeth width and from dimension Δs.

- Synchronizer body width must be such that the sleeve does not engage any synchronizer ring in the neutral position.

12.2.2 Functional criteria

If we assume that the apparent rotating mass of the vehicle is much greater than that of the parts to be synchronized, the synchronization transient will resemble that of the clutch during vehicle start-up and will be described with a single equation, because with this assumption the speed of the vehicle will remain constant.

If we designate as M_s the synchronization torque, as M_a the friction torque of the mechanical losses of the parts to be synchronized, as Ω the rotation speed of elements upstream of the synchronizer, and as $J_{r,eq}$ the rotating mass of the parts to be synchronized, we can write:

$$M_s + M_a = -J_{r_{eq}} \frac{d\Omega}{dt} \ . \tag{12.6}$$

Because M_s and M_a are almost constant with speed, angular acceleration will be constant in time while rotation speed will decrease constantly in time; during up shifts the acceleration is negative, while during down shifts it will be positive.

Synchronization torque will always show a sign opposite to that of angular acceleration, while friction torque will show the same sign as synchronization torque during up shifts, when the rotating masses must be slowed.

In the above hypotheses thermal power $P_{\max}$ and thermal energy E wasted by the synchronizer are:

$$P_{\max} = M_s \left(\Omega_2 - \Omega_1\right) \ , \tag{12.7}$$

$$E = \frac{1}{2} J_{r_{eq}} (\Omega_2^2 - \Omega_1^2) + \frac{1}{2} M_a \left(\Omega_2 - \Omega_1\right) \Delta t \ , \tag{12.8}$$

where subscripts 1 and 2 apply to initial and final conditions.

Let us now assume that we have to design the first gear synchronizer of a two stage gearbox, in terms of the second to first shift; the scheme of the gearbox is shown in Fig. 12.9.

We designate as z_i the gear teeth numbers (i refers to the gear number in the diagram) and as J_i the related rotating mass. The equivalent rotating mass $J_{r,eq}$ to put in the above equation is that of all rotating masses upstream the synchronizer, reduced to the gear 3:

$$\begin{aligned} J_{r_{eq}} = J_3 + (J_d + J_1) \left(\frac{z_3}{z_8}\right)^2 \left(\frac{z_7}{z_1}\right)^2 + (J_{ca} + J_7 + J_8 + J_9) \left(\frac{z_3}{z_8}\right)^2 + \\ + J_2 \left(\frac{z_3}{z_8}\right)^2 \left(\frac{z_8}{z_2}\right)^2 + J_4 \left(\frac{z_3}{z_8}\right)^2 \left(\frac{z_9}{z_4}\right)^2 \ . \end{aligned} \tag{12.9}$$

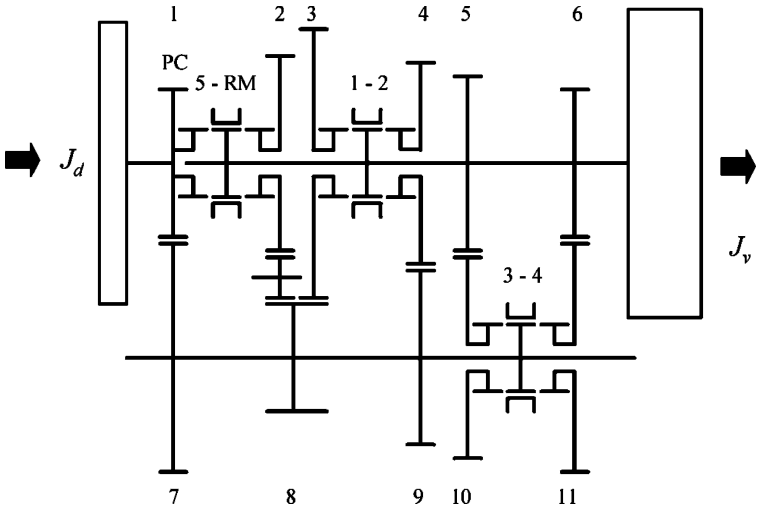


FIGURE 12.9. Scheme for the calculation of the equivalent rotating mass to be synchronized during a gear shift manoeuvre.

The masses of synchronizers and reverse idlers are neglected; J_{ca} is the countershaft's rotating mass. Idle wheels on the countershaft must not be taken into account, because they are already synchronous with the output speed.

On the same gearbox, if the synchronizer is put on the countershaft, the rotating mass will be:

$$J_{r_{eq}} = J_8 + J_{ca} + J_9 + (J_d + J_1)\left(\frac{z_7}{z_1}\right)^2 + J_2\left(\frac{z_8}{z_2}\right)^2 \ , \tag{12.10}$$

because the remaining wheels will be rotating with the output shaft.

The advantage of this second configuration is apparent; it can also be noticed that the rotating mass decreases notably while the speed order increases. On a single stage gearbox the calculation will be simpler and will confirm the above conclusions.

After the evaluation rotating masses it is possible to calculate synchronization time as a function of the shifting force; in the same way wasted power can be calculated and compared, as a design parameter, with other successful gearboxes.

The synchronizer reference speed is defined by:

$$V = \frac{d}{2}\left(\Omega_2 - \Omega_1\right) \ , \tag{12.11}$$

showing the relative speed at the synchronizer reference diameter at the beginning of the shift manoeuvre.

The empirical reference values, reported in Table 12.1, can be considered for the above magnitudes.

Specific power and energy are derived from wasted power and energy by dividing by the active surface of the synchronizer.

Because as high as possible a friction coefficient is wanted, a lubricated hydrodynamic film should be avoided.

Nevertheless, synchronizers are splashed with lubricating oil; with reference to what we said above, the active surface of synchronizers is grooved or threaded; threads should unwind at the relative speed.

If surfaces are grooved or threaded the reference surface area must consequently be reduced.

Something more must be said about the tapered ends of the dog teeth; tangential force consequent to the opening must not rotate the ring before synchronization has been accomplished.

If we represent some of these teeth in Fig. 12.10, we can see that, under the action of the shifting force F, the opening angle of the end sets up a torque

TABLE 12.1. Reference design values for synchronizers made by steel/brass and steel/steel with molybdenum coating.

Material	f	V (m/s)	E_s (J/mm^2)	P_s (W/mm^2)	p (N/mm^2)
St./Br.	$\backsim$0.1	5	0.09	0.45	3
St./St,m.	$\backsim$0.1	7	0.53	0.84	6

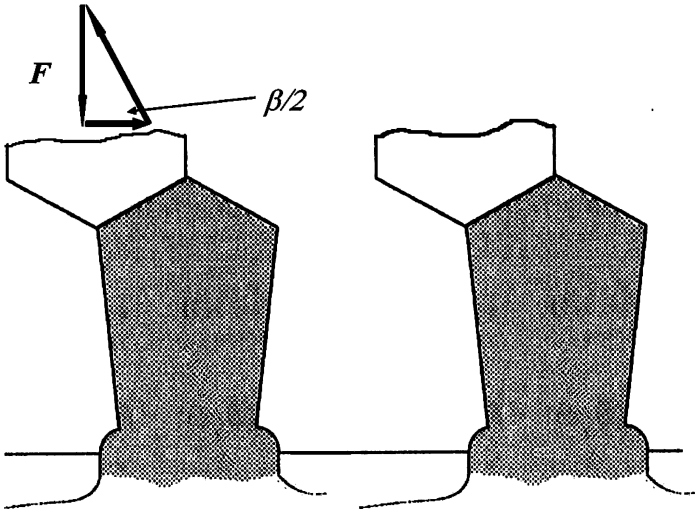


FIGURE 12.10. Scheme for the calculation of the condition of possible engagement before the accomplishment of the synchronization.

of opposite direction to that of synchronization; if the first were greater than the second, the shifting force would be strong enough to engage the dog clutch before synchronization is accomplished.

If β is the opening angle of the tooth and d_s is the mean diameter of the teeth, the torque acting on the ring will be:

$$\frac{1}{2} d_s F \cot\left(\frac{\beta}{2}\right) ,$$

if we neglect friction on the end of the tooth; this torque must not be greater then the synchronization torque:

$$\frac{d_f F}{2 \sin(\alpha)} .$$

Matching materials for synchronization ring and hub must proceed by taking into account the following needs:

- Wear must be low without the help of hydrodynamic lubrication.
- Materials must be easily machined.
- The friction coefficient must be as constant as possible in production.
- The friction coefficient must be insensitive to wear and temperature.
- Materials must withstand possible overloads.

As a matter of fact the available options are few; hubs are usually made of Cr Mn steel or Cr Mo steel. Rings can be made of special shot peened brasses or steel, coated with a molybdenum layer. A paper layer could also be considered in the future, with advantages for the friction coefficient.

13

DIFFERENTIALS AND FINAL DRIVES

The terms differential, transfer box and final drive are sometimes used imprecisely; in addition, these mechanisms are often integrated in the same subsystem in a wide variety of combinations. In this chapter we introduce definitions that should help to clarify this issue.

The *differential* is a mechanism that allows the torque from an input shaft to be divided into two predetermined parts, flowing through two output shafts; torque ratios are independent of the speed ratios of the same shafts.

This mechanism can be used either to divide the torque coming out of the final drive into equal parts acting on the traction wheels of the same axle, or to divide the torque coming out of the gearbox into two predetermined parts acting on different axles of the same vehicle. This second application is sometimes called the *transfer box differential* or *central differential*.

The *final drive* is a gear train that further reduces the speed of the gearbox output shaft to adapt it to the traction wheels; this gear train is usually integrated with a differential mechanism. Under this name are sometimes included those speed reducers that are put on the transmission line after the differential final drive and integrated in the wheel hub.

The *transfer box* is a mechanism that provides for the movement of two or more drivelines through the single output shaft of the gearbox; it is used on vehicles with multiple traction axles. When multiple axle traction is permanent a differential is also needed to allow different mean rotation speeds on the axles; in such cases the differential train is usually integrated in the transfer box.

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

13.1 DIFFERENTIALS AND FINAL DRIVES

Different schemes exist according to the type of vehicle drive; in all of them the differential mechanism usually shows the same configuration.

Look at the diagrams shown in Figs. 1–3 to identify all possible configurations.

13.1.1 Rear wheel driven cars

On rear wheel driven cars final drives carry out the task of rotating the driveline axis 90 deg, from the longitudinal gearbox center line, to the transversal axle center line. In Fig. 13.1 the cross section of this subsystem is shown. Center line rotation and speed reduction are achieved through a pair of bevel gears with spiral teeth; the wheel is bolted to a hub in order to allow the easy adaptation of the same production line to different final transmission ratios.

Shafts are supported by bevel roller bearings, because the axial thrust is relevant. The differential train is made with straight teeth bevel gears, because their rotation is not continuous and their speed is low. *Planetary wheels* are fixed to half axles through splines, while *satellites* are idle on a short shaft fixed to a carrier through a pin; the bevel gear wheel of the final drive is bolted onto this carrier.

The differential assembly is a suspended, used in connection with independent suspensions: Half axles will therefore be moved through constant speed

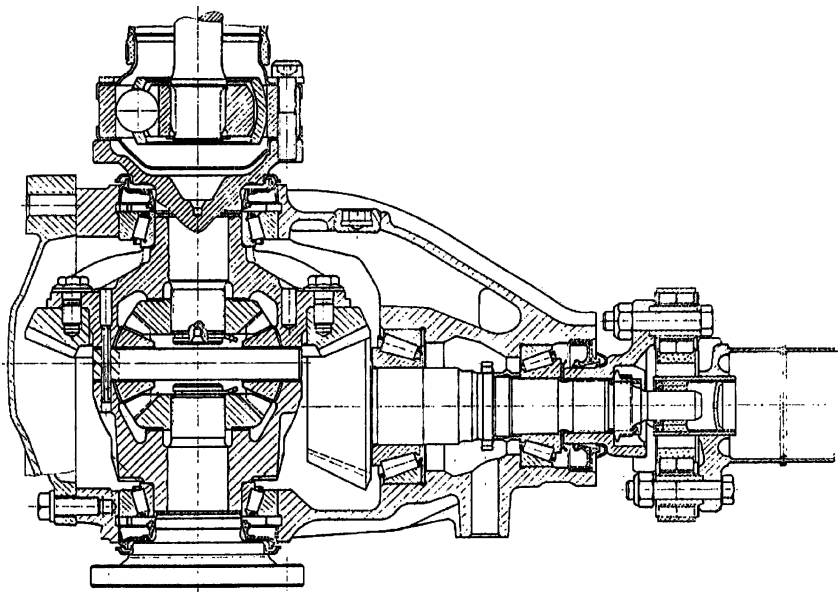


FIGURE 13.1. Cross section of a differential assembly for a rear wheel driven car with rear independent suspensions (Mercedes).

joints, as in a front wheel driven car. It should be remembered that a suspended differential casing must react to the vectorial resultant of the wheel traction torque, acting on the xz plane, and of the gearbox output torque, acting on the yz plane; for this reason the casing features a robust fixation interface, to suspend it from the car body, usually through an auxiliary frame.

Notice in the same figure the small elastic tube on the bevel pinion shaft; it allows the axial pre-load to be controlled on bevel roller bearings, compensating for axial dimension tolerance.

In this application the final drive bevel gears feature *hypoid* teeth, because their centre lines lie on different planes; this configuration allows the transmission line to be placed in a lower position with advantages for interior roominess.

Planetary wheels and satellite thrust bearings are simple wear resistant washers, because of the low relative speed.

13.1.2 Front wheel driven cars

On front wheel driven cars differential and final drive are integrated in the gearbox; the reaction torque therefore acts on the power train suspension. We can identify two cases where the engine is transversal or longitudinal.

In the first case, shown in Fig. 13.2, the gearbox output shaft and wheel centre lines are parallel; a pair of spur gears with helical teeth is sufficient. The pinion is cut on the output shaft and meshes with a wheel, which in this case is also flanged to the differential carrier.

The differential mechanism is similar to the previous. In this case planet wheel thrust bearings are made with needle cages to improve mechanical efficiency. A following paragraph will explain the influence of this mechanical efficiency on the dynamic behavior of the vehicle.

For the longitudinal engine, shown in Fig. 9.9, there are no relevant differences in comparison with rear wheel drive; hypoid teeth are usually unnecessary. In Fig. 13.3 we see the differential assembly corresponding to the traction scheme in Fig. 1b; the bevel pinion is fixed on a very short transmission shaft that allows the front axle to be shifted in front of the gearbox output shaft end.

The driving wheel of this short shaft is a bevel gear and a simple damper for flexural vibration has been provided. This kind of architecture is necessary when front wheel overhang must be limited or when automatic epicycloidal gearboxes are used, where input and output shafts are aligned.

13.1.3 Industrial vehicles

On industrial vehicles the final drive and differential train are integrated in the same rigid structure supporting wheels and half axles, called the rigid axle (see Fig. 13.4).

Notwithstanding the increased robustness of the group, there are no noticeable differences in comparison with rear wheel driven cars: The final drive is

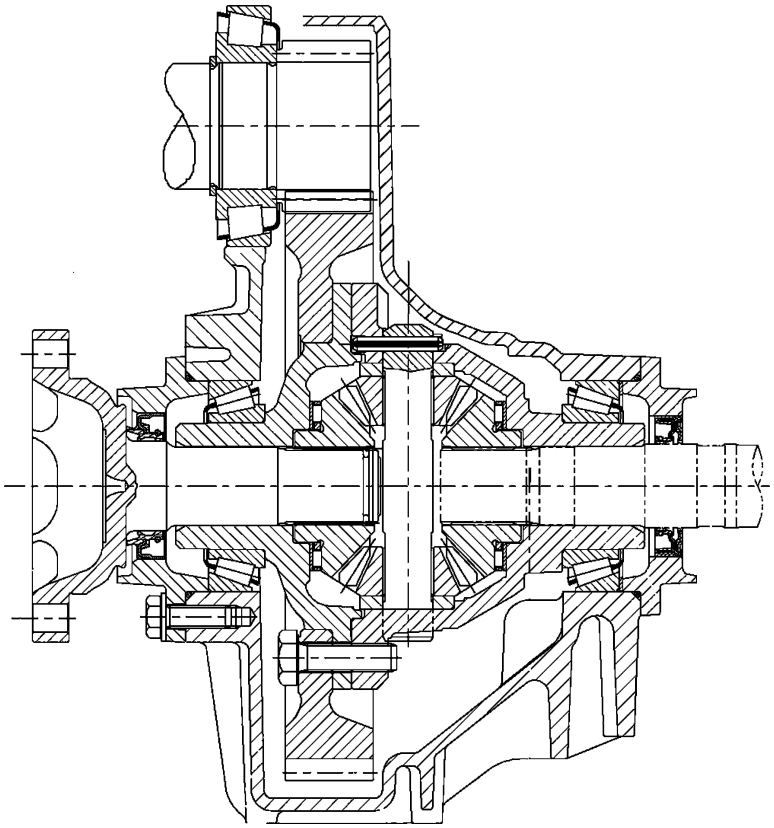


FIGURE 13.2. Differential and final drive for a front wheel driven car with transversal engine (FIAT).

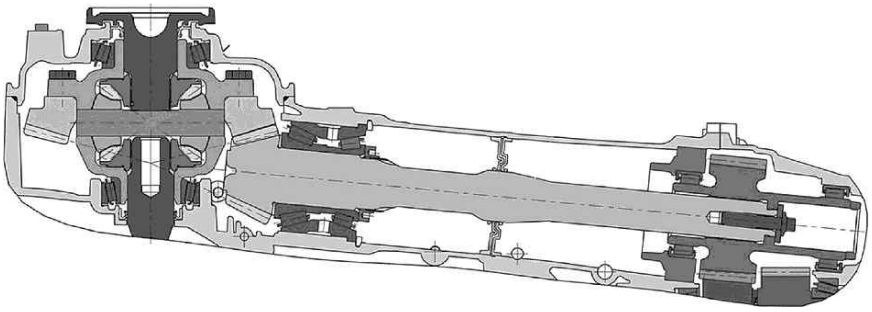


FIGURE 13.3. Differential and final drive for a front wheel driven car with longitudinal engine; in this case the wheel axis is shifted forward in order to reduce front overhang (Audi).

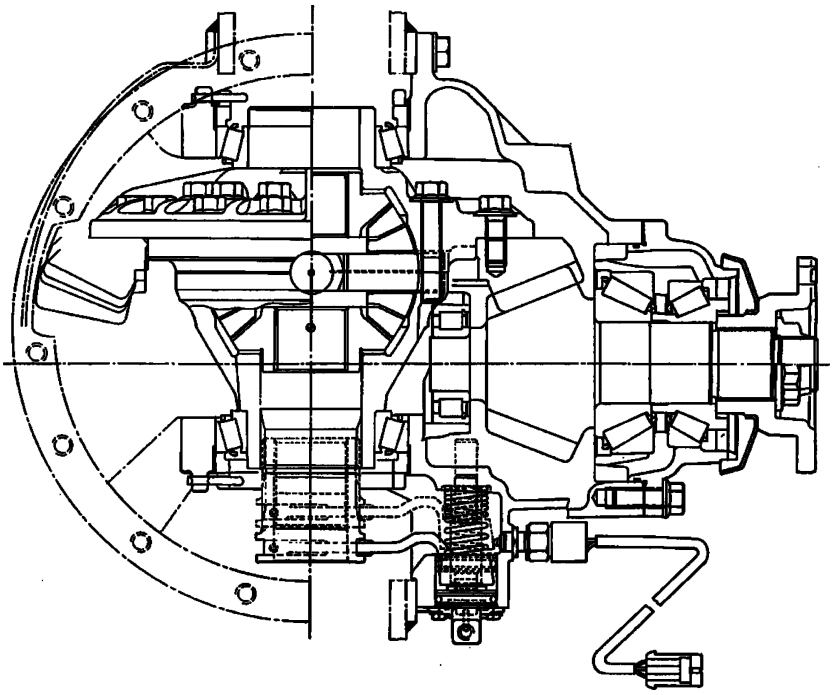


FIGURE 13.4. Rigid axle differential assembly for a heavy duty truck (Iveco). The differential lock dog clutch can be seen on the right half axle.

made with bevel gears with hypoid teeth. Wheel offset is used in this case in the opposite way as before, to increase transmission shaft ground clearance.

On the right half axle a locking sleeve can be noticed; the lock is particularly useful to improve start-up capability on a slippery road.

For construction or off-road vehicles transmission ratios obtained with a single stage final drive may not be sufficient; transmission ratio is limited by the allowed bevel gear dimension and by the minimum number of teeth that can be cut on the pinion once the pitch has been defined.

This problem can be solved by installing an additional epicycloidal final drive. In this case the differential carrier can be integrated with the epicycloidal drive carrier, where the annulus gear is fixed on the bevel wheel shaft; in this way, the additional drive does not change the direction of rotation.

On urban buses is quite important to limit the ground height of the aisle floor. Floor height and tire rolling radius define the maximum diameter of bevel gear wheel; therefore, as represented in Fig. 13.5, a spur gear final drive is necessary on the output shaft of the differential. This choice is motivated by the need to have a bridge shaped rigid axle whose central part is compatible with a low aisle floor.

The inverted bridge configuration can be also used to increase ground clearance on off-road vehicles; in this case will hubs will be lower than the half axles.

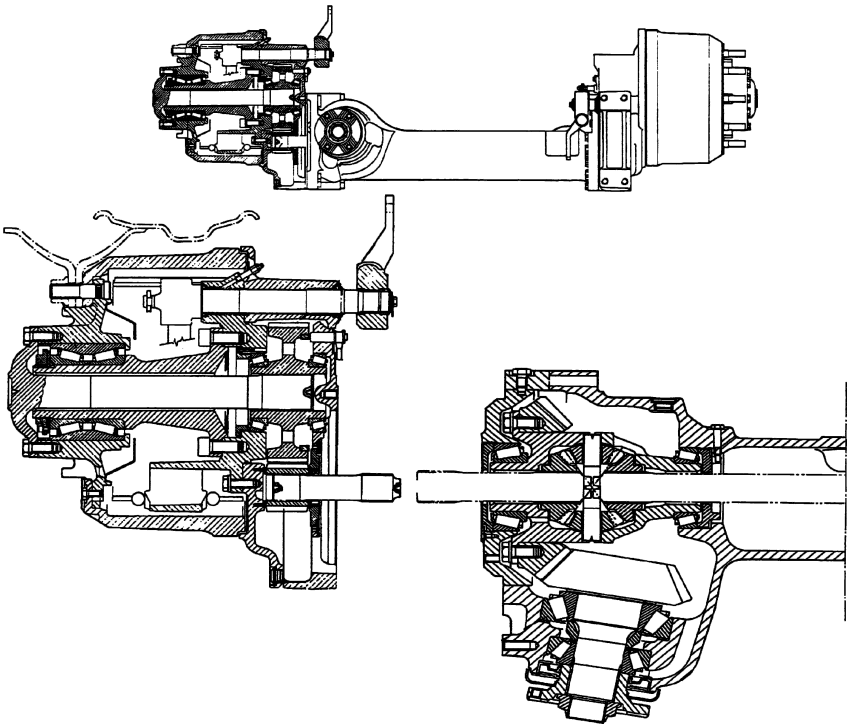


FIGURE 13.5. Rigid rear axle for urban buses (Iveco). The lower detail shows the additional final drives on the wheel hubs that shape the axle like a bridge. Note the inclination of the input shaft, caused by the transversal power train lay-out.

13.2 ALL WHEEL DRIVE TRANSFER BOXES

On all wheel driven vehicles transfer box architecture is conditioned by two factors:

- The vehicle single axle traction configuration the all wheel drive is derived from, with reference to the architectures shown in Fig. 3.

- The fact that the drive is *permanent* (can be used at any vehicle speed) or *non-permanent* (can be used occasionally at low speed on bad roads). In the first case a central transfer differential is also necessary, to avoid unnecessary tire wear and additional rolling resistance when wheel speeds differ; in the second case the differential can be avoided.

We consider drivelines derived from conventional rear wheel driven vehicles and front wheel driven vehicles, either with transversal or longitudinal engine.

13.2.1 Modified rear wheel drives

The architectures in use are many; a possible classification can be derived from the vehicle use. We can identify off-road vehicles, where the main objective is to obtain good mobility on dirty and slippery roads, and road vehicles, where the objective is to achieve superior stability and handling on paved roads at high speed.

In off-road vehicles (scheme c, Fig. 3), the transmission line is offset in comparison to the engine crankshaft to increase ground clearance. The transfer box works for rear axle drive only; the transfer box lay-out is visible in Fig. 13.6.

Other functions are present; as a matter of fact the gearbox output shaft moves a synchronized two speed range reducer. The reducer output shaft moves a third shaft featuring a bevel gear differential that allows the torque to be split between the two axles. In the same figure a dog clutch is also shown; it can fix the differential carrier with the front planet gear, in this way locking the differential for slippery roads at low speed.

The offset between crankshaft and transmission line is useful because it allows rigid axle suspensions to be used without interference with the oil sump.

This type of transfer box can have a simple dog clutch instead of the differential to insert the front wheel drive on low speed slippery road.

On on-road vehicles, the original rear wheel drive line position is conserved; a transfer box is added to move the front axle, using a central differential at all times.

An example is shown in Fig. 13.7, where motion to the front axle is given through an idler. The differential is of the epicycloidal spur gears type, where the carrier is fixed to the rear axle transmission shaft, the sun gear moves the front drive transmission shaft, and the annulus gear is driven by the gearbox output shaft.

A multi-disc wet clutch controls the two axle speed difference; we will comment later on the function of this clutch.

A second clutch, at the right of the figure, can put the front axle in idle. Three operating modes are available: Rear wheel drive, constant rate all wheel drive, locked differential all wheel drive.

It should be noticed that the carrier shows twin satellites, to allow the correct rotation direction for the front axle.

The use of a spur gear epicycloidal differential allows any torque breakdown rate other than 50/50; this is useful when axle loads are different and, in any case, correctly controls the vehicle static margin, important for vehicle stability.

13.2.2 Modified front wheel drive vehicles

It may be useful, to distinguish between transversal and longitudinal engines.

In the most widely used types of transversal front engine, shown in Fig. 13.8, the front axle differential is coupled to a simple bevel gear drive featuring a transmission ratio close to one. The driving bevel gear is fixed to the front axle

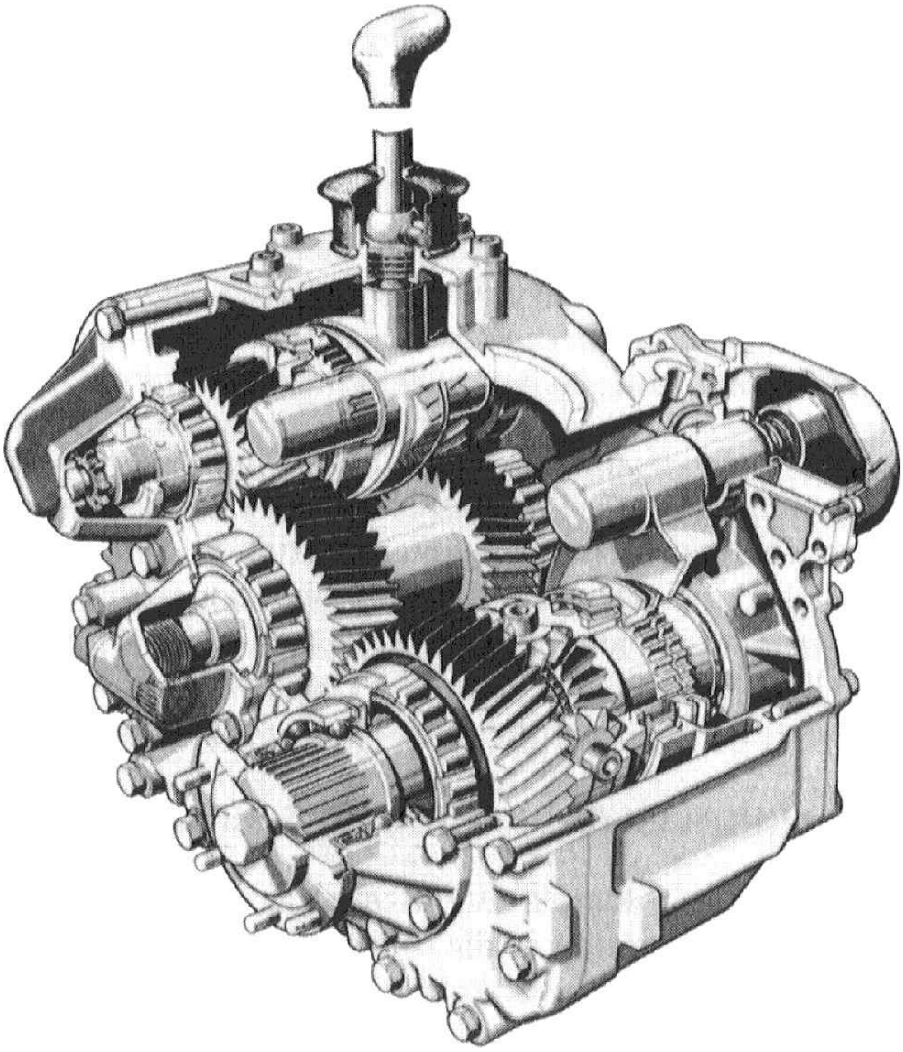


FIGURE 13.6. Transfer box for a professional off-road vehicle (Mercedes). The transfer box integrates the two axle drive through a bevel gear lockable differential and a range speed reducer with two synchronized speeds.

differential carrier; the rear axle transmission shaft is moved by the driven bevel gear. There can be a simple dog clutch, not shown in this figure, to move the rear axle, in case of a non-permanent all wheel drive.

The rear differential group is similar to that of a rear traction drive; in case of permanent traction, a viscous coupling is usually fit on the rear axle drive line. This joint allows part of the torque available to the front axle to be subtracted, only when this axle shows an average speed greater than that of the rear axle. In a following paragraph we will comment about this joint and its working method.

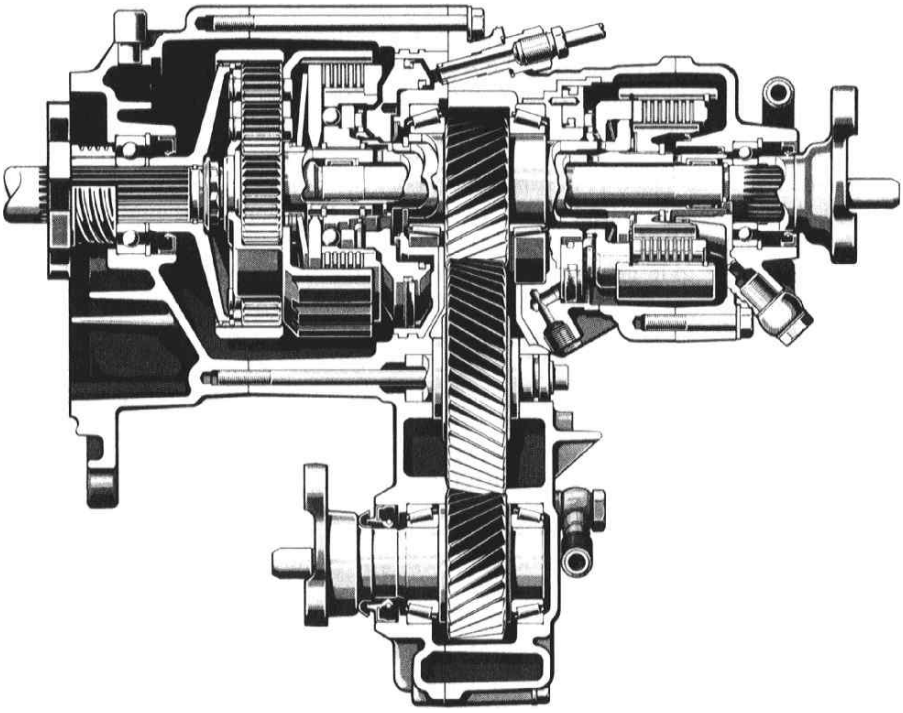


FIGURE 13.7. Transfer box for an all wheel drive on-road vehicle with longitudinal front engine (Mercedes). The box integrates the axle drive with an epicyclical differential with slip control.

A self-locking transfer differential can be installed in the transfer case; here the viscous coupling is avoided.

The longitudinal engine is shown in Fig. 13.9. All wheel drive looks simpler; the gearbox output shaft is hollow. The shaft moving the front differential bevel pinion is left free to turn in the cavity. The gearbox output shaft moves a Torsen type differential, as we will explain later. The planet wheels of this differential move respectively the front and the rear axle. This differential transfers and controls the torque to the two axles.

13.3 OUTLINE OF DIFFERENTIAL THEORY

In these paragraphs the influence of differential mechanical efficiency on the value of transferred torque is examined and compared with the case of an ideal friction-free differential.

The friction torque is sometime deliberately increased with different devices to control the value of the output torque. These devices will be examined and explained and their effect on vehicle dynamics will be investigated.

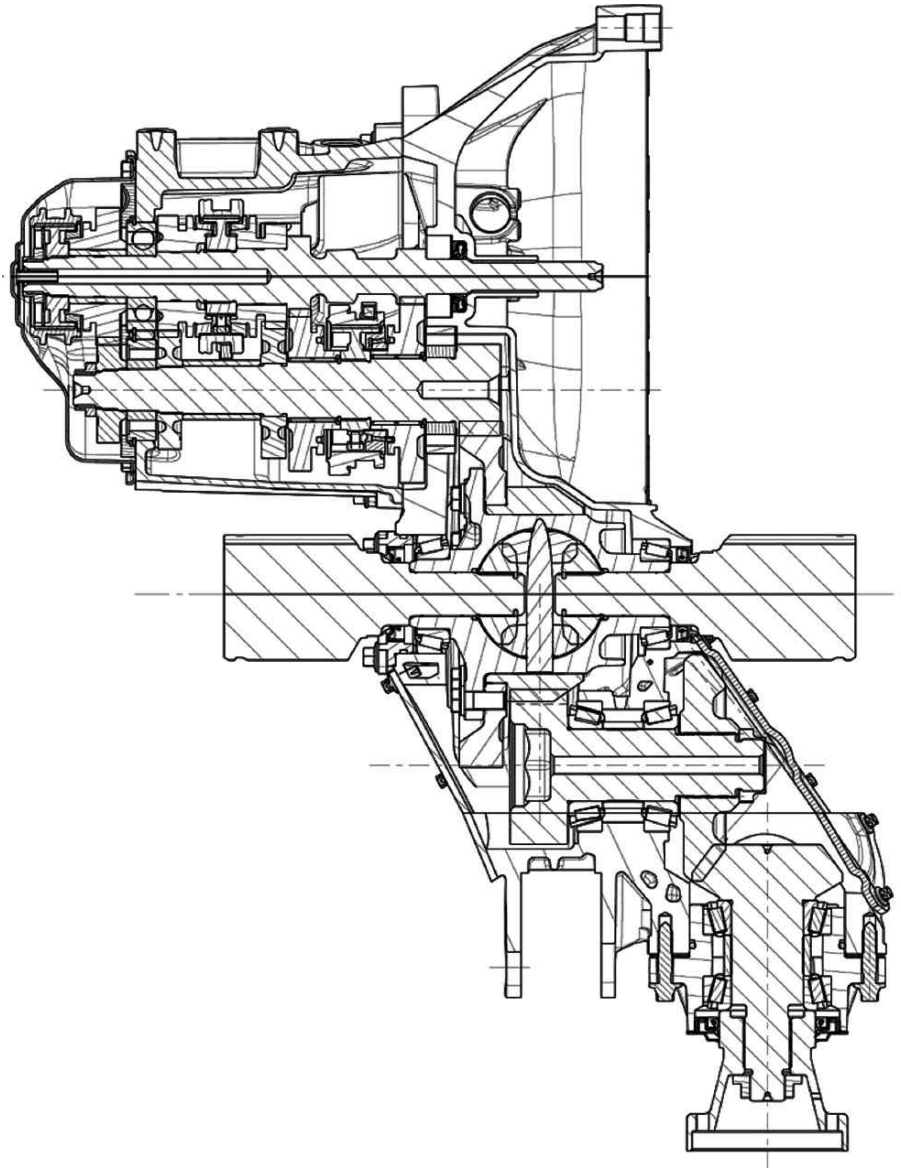


FIGURE 13.8. Gearbox, differential and transfer box of an all wheel drive car with transversal front engine (FIAT).

13.3.1 Friction free differential

The differential is a two degree of freedom mechanism that can be idealized as a black box where an input shaft enters at a speed Ω and with a torque M; from this black box, two output shafts exit at Ω_1 and Ω_2 speeds, with M_1 and M_2

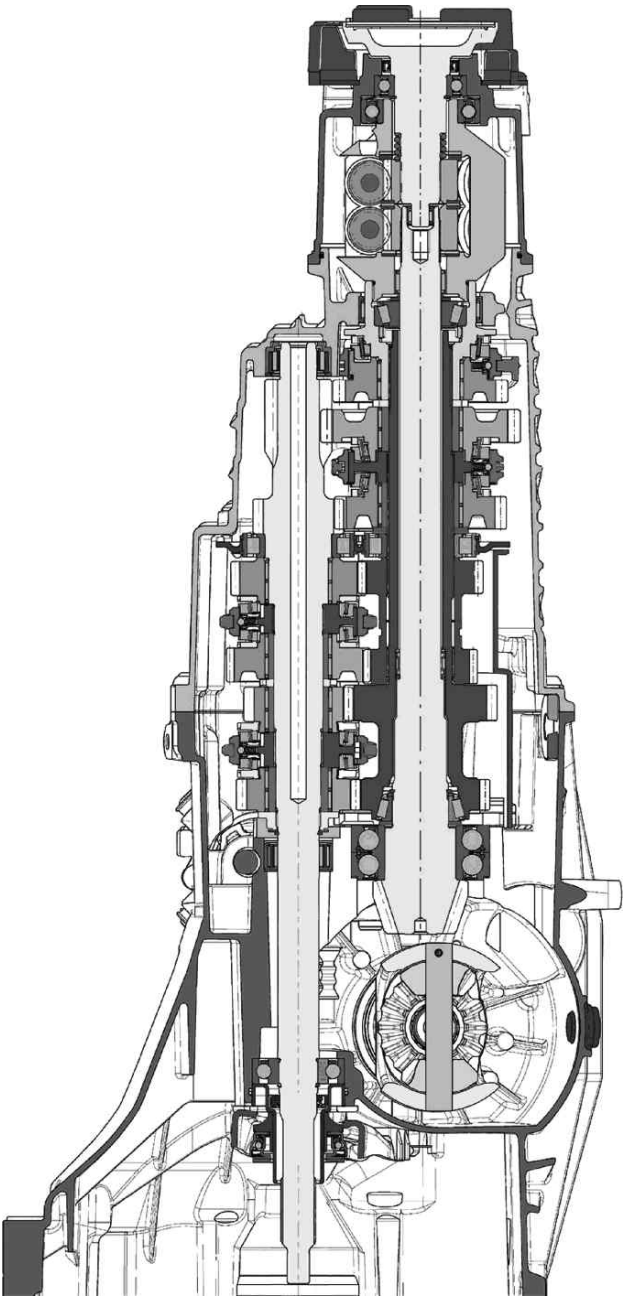


FIGURE 13.9. Gearbox, differential and transfer box of a permanent all wheel drive with longitudinal front engine (Audi). The gearbox output shaft is hollow, in order to install the front wheel axle drive shaft.

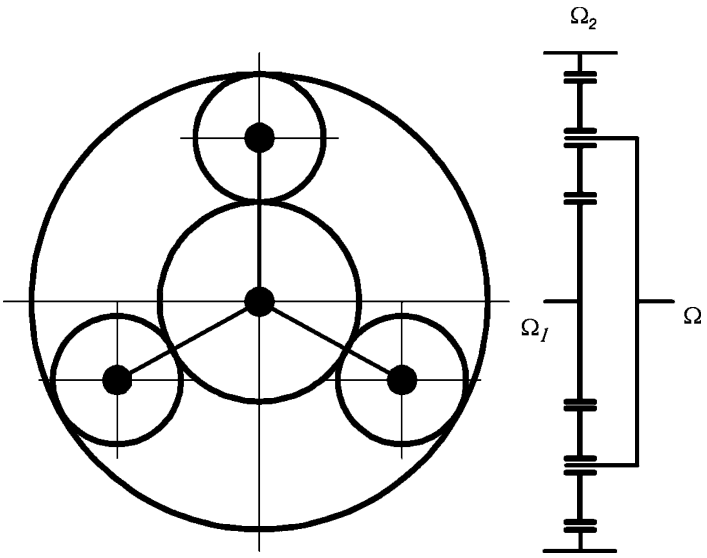


FIGURE 13.10. Kinematic scheme of an epicyclical train, used as a differential.

torques. The two degrees of freedom could be the input shaft rotation angle and the difference in angular displacement of the two output shafts.

The properties of a differential must be as follows:

1. There must be only one relationship between the three speeds; in this way the speed difference between the output shafts is undetermined (the mechanism is named after this property).

2. The input torque is split into two output torques, in a constant ratio independent of speed.

3. The two output torques must have the same direction.

These properties are fulfilled by an epicycloidal gear such as that sketched, with the usual rules, in Fig. 13.10. If r is the ordinary transmission ratio, i.e. the transmission ratio of the one degree of freedom mechanism, obtained by locking the carrier, we can write:

$$r = \frac{\Omega_2 - \Omega}{\Omega_1 - \Omega} \ , \tag{13.1}$$

and:

$$\Omega_2 = \Omega + r(\Omega_1 - \Omega) \ , \tag{13.2}$$

or:

$$\Omega_2 - \Omega_1 = (1 - r)\,(\Omega - \Omega_1) \ . \tag{13.3}$$

From the last formula we see that the speed difference of the two output shafts can assume any value. The sign of r is negative, because when the carrier is locked output speeds are opposite.

For the system equilibrium it must be true that:

$$M_1 + M_2 = M \ ; \tag{13.4}$$

if there is no friction it must also be that:

$$M_1\Omega_1 + M_2\Omega_2 = M\Omega \ . \tag{13.5}$$

If we substitute one of the speed equations in the last, we obtain:

$$M_1 = - \ M \ \frac{r}{1-r} \ , \tag{13.6}$$

$$M_2 = M \ \frac{1}{1-r} \ , \tag{13.7}$$

$$\frac{M_1}{M_2} = -r \ , \tag{13.8}$$

where we can see that the torque is split into constant parts, independent of speed.

The well known case of conventional bevel gear differentials can be described by the above equations where the ordinary transmission ratio is set to -1; in this case, output torques are always equal to half the input torque.

In a transfer box differential, if a split ratio other than 50/50 is desired, it is necessary to use an ordinary transmission ratio other than -1.

13.3.2 Differential with internal friction

Let us assume now that mechanical losses do exist and that they can be expressed through the concept of mechanical efficiency, which states that the power loss is proportional to useful power.

We define η as the mechanical efficiency of the ordinary gear train associated with the differential; if the shaft 2 rotates in the same direction as the torque M_2, this torque will be called input torque and M_1 will therefore be output torque; under this assumption:

$$M_1 = -M_2 r\eta \ ; \tag{13.9}$$

if, instead M_1 is input and M_2 is output, it will be:

$$M_2 = -\frac{M_1\eta}{r} \ . \tag{13.10}$$

The efficiency η does not change value, as a first approximation, under the two different directions of power flow; if Ω is different than zero, the first formula is valid when the shaft 2 is slower and the second formula when the shaft 2 is faster.

Therefore, if $\Omega_1 > \Omega > \Omega_2$ it will be:

$$M_1 = \frac{r\eta}{1-r\eta}, \tag{13.11}$$

$$M_2 = M\frac{1}{1-r\eta}, \tag{13.12}$$

$$\frac{M_2}{M_1} = -\frac{1}{r\eta}. \tag{13.13}$$

Instead, if $\Omega_1 < \Omega < \Omega_2$ it will be:

$$M_1 = -M\frac{r}{\eta - r}, \tag{13.14}$$

$$M_2 = M\frac{\eta}{\eta - r}, \tag{13.15}$$

$$\frac{M_1}{M_2} = -\frac{r}{\eta}. \tag{13.16}$$

If the three speeds are equal, the output torque will be undetermined within a range included between the values of the above formulae, analogous to the friction phenomena, where there is no relative motion, if friction torque (or force) is still below the static friction value. The case of the axle differential can again be obtained by setting $r = -1$.

Contrary to the ideal case of no friction, if one wheel can bear a lower traction force because of a lower traction coefficient, there will be no relative motion between the wheels if the difference in torque is included in the above limits.

In the case of a bevel gear differential, we can refer to the scheme in Fig. 13.11. Forces exchanged by planet wheels and satellite teeth are shown. The scheme represents the three reference planes of the bevel gears as if they were on the same plane; planet gear drawing planes are placed over those of satellites.

On the flank of the satellite tooth flank tangent forces T_2 and T_1 will act, inclined according to the pressure angle ϑ.

But, because of the friction, these forces will also cause normal components N_1 and N_2, which must be oriented in such a way as to oppose the relative

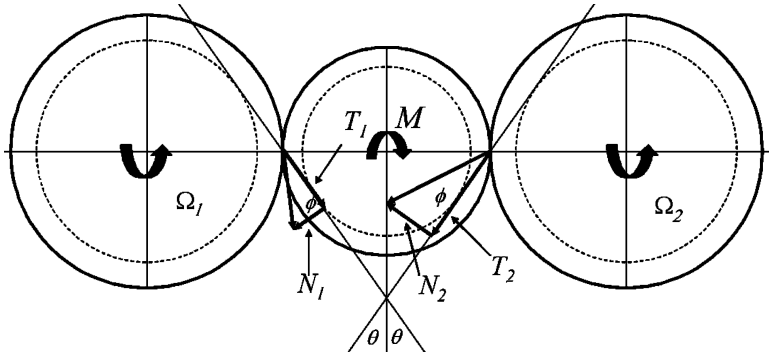


FIGURE 13.11. Scheme of forces acting on a bevel gear differential; reference planes of planet wheels are represented coincident with the reference planes of satellites.

motion; in this scheme we assume that:

$$\Omega_1 > \Omega > \Omega_2 \ .$$

If R is the reference radius of the primitive satellite cone, we can write the rotation equilibrium equation of satellites, bearing in mind that the friction angle ϕ is bound to the friction coefficient f; it must be that:

$$\begin{aligned} & f = \tan(\phi) \ , \\ & T_1 \tan(\phi) = N_1 \ , \\ & T_2 \tan(\phi) = N_2 \ , \\ & (T_2 - T_1) R \cos(\vartheta) = T_1 + T_2 \ . \end{aligned} \tag{13.17}$$

If we designate with R' the reference diameter of the planet wheel primitive cone, it must be that:

$$(T_1 + T_2) R' = M \ . \tag{13.18}$$

Therefore, we obtain the final equation:

$$\eta = \frac{T_1 R'}{T_2 R'} = \frac{1 - f \tan(\vartheta)}{1 + f \tan(\vartheta)} \ . \tag{13.19}$$

13.3.3 Self-locking differential

A *self-locking* differential is a particular mechanism where the torque difference between the two output shafts can be limited to preset values; the limit case is when the differential is completely locked where the speed torque is undetermined (planet wheels are locked together).

We define as *locking coefficient* b of a differential the maximum difference of torque (referred to the total axle torque) between the two output shafts that can be sustained with no difference in speed; the torque difference is called *locking torque*.

We have by definition that:

$$b = (M_2 - M_1)/(M_2 + M_1) \ , \tag{13.20}$$

or:

$$\eta = (1 - b)/(1 + b) \ . \tag{13.21}$$

The b coefficient can be constant or can depend upon speed difference or input torque values; we will see some practical examples in the next paragraph.

We remember finally that the name self-locking differential is given to devices where the locking torque is determined by the mechanical properties of the system; the name *controlled differential* is given instead to devices where the locking torque is determined by a controlled clutch using, for example, electronic means.

13.4 TYPES OF SELF-LOCKING DIFFERENTIALS

13.4.1 ZF system

In the ZF system, shown in Fig. 13.12, the bearings where the thrust of the planet wheels 6 is loaded are not made by anti-friction material, but by a multi-disc wet clutch.

In the above figure, two discs 5 rotate with planet gears, while two other discs 4 and the pressure plate 3 rotate with the carrier. Pressure plates 3 show a V shaped groove (detail a) that matches with satellite shafts 3 where total input torque is applied.

By means of this device, the discs pressure force is about proportional to the total torque. The locking torque is therefore proportional to the total torque and will always oppose the relative rotation of the half axles. The locking coefficient b is constant.

On the lower part of the same figure, a different version of this differential is presented; the pressure plates are pushed by a diaphragm spring 7. In this case the locking torque features a minimum locking value also at zero total torque; the blocking coefficient won't be constant anymore.

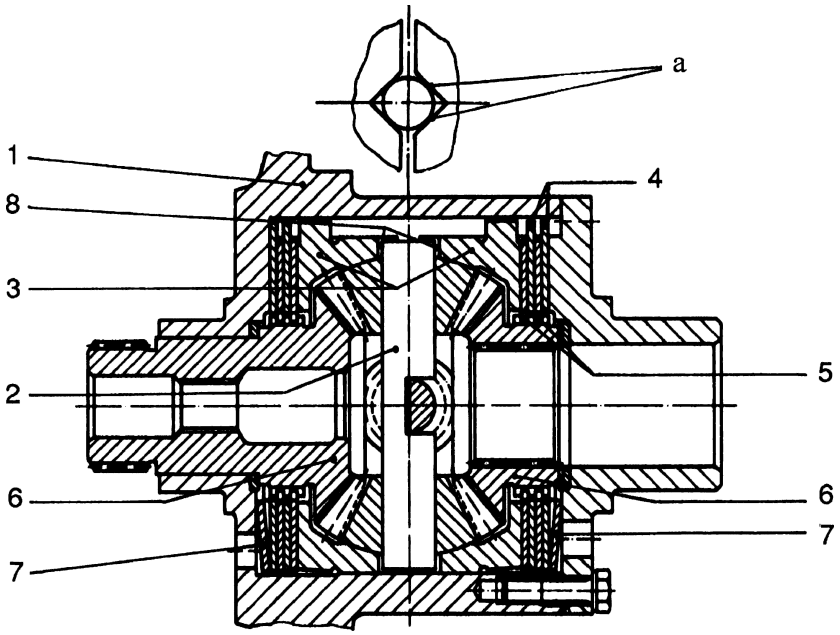


FIGURE 13.12. Self-locking ZF differential. The multi disc wet clutches apply a constant (upper version) or a variable (lower version) locking coefficient.

13.4.2 Torsen system

A cross section of a Torsen differential (Torsen stands for torque sensitive) is represented in Fig. 13.13; planet wheels 5 and satellites 3 are made with worm gears. The centre line of the satellites lies on a plane that is perpendicular to the centre line plane of the planet gears.

There are three satellites for each planet gear. The satellite teeth do not mesh with those of the neighboring satellite; small spur gear wheels 4 can counter rotate each pair of neighboring satellites.

The satellite shafts are blocked in their radial and axial direction; only their rotation is free. It should be noted that, in order for the mechanism to work correctly, the wheel threads must have the same direction.

Under these conditions, the ordinary transmission ratio is again –1; as a matter of fact, if we lock the carrier 2, a planetary rotation will cause the corresponding rotation of the satellite that will be reversed in the neighboring satellite which, finally, will cause the other planet wheel to rotate in the opposite direction, at the same speed.

The relevant friction existing on helical surfaces creates poor mechanical efficiency; a simplified mathematical model can lead to the same conclusions as those of bevel gear differential. In this case also, the blocking coefficient will be constant.

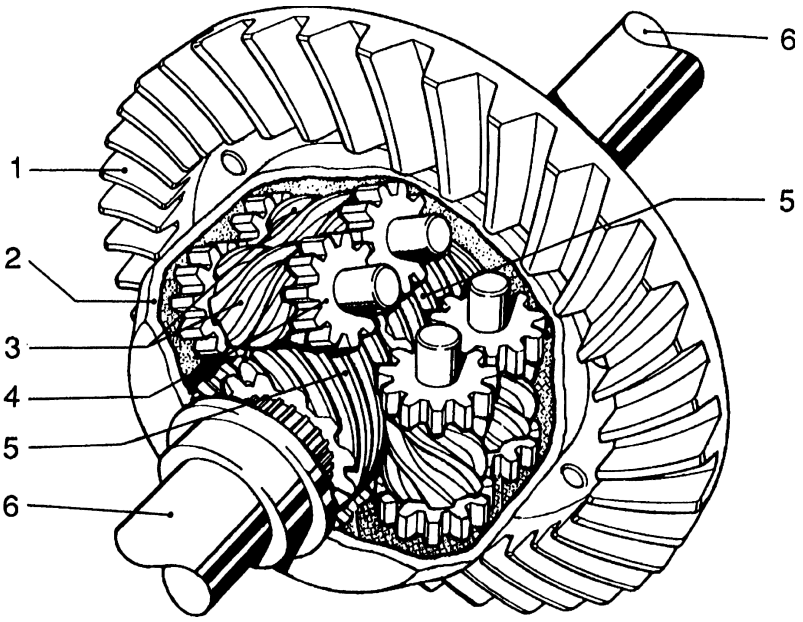


FIGURE 13.13. Scheme of a Torsen differential. The locking torque is due to the high friction existing between the worm gear planet wheels and satellites.

13.4.3 Ferguson system

The Ferguson system is made by applying a viscous joint to two different elements of a differential; the joint exploits the oil (silicon oil) viscosity to brake two elements of the differential with a torque depending on the relative speed of the elements.

Figure 13.14 shows the scheme of a joint suitable for the transmission shaft of an all wheel drive gearbox similar to that in Fig. 13.8; this joint can also be integrated in a differential according to different schemes.

The joint is made by a cylindrical reservoir 3 fixed to shaft 2, on the right of the figure. A second shaft 1, on the left, is free to rotate in the reservoir; a rotating seal avoids oil leakage outside the reservoir.

The left shaft has a spline where a series of metallic discs 6 is invested; another series of discs 5 is connected to the inside diameter of the reservoir, again through a spline. The two disc families are stacked with each of the discs 6 inserted between two discs 5.

The discs are drilled or cut to activate viscous forces between the disc families, when there is a relative rotation speed; axial forces are unnecessary, because the braking force is controlled by the oil speed gradient in gaps between facing discs. The torque value is determined by disc number (increasing with the number) and by the apparent oil density; this last can be modified easily by changing the quantity of air emulsified in the oil (decreasing with the quantity of air); these two magnitudes, together with disc diameter, are the design parameters of the joint.

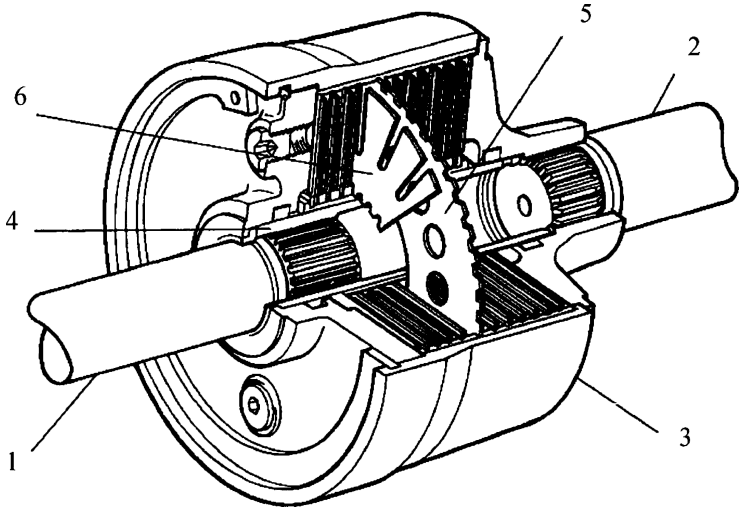


FIGURE 13.14. Scheme of a Ferguson joint. The friction between the two shafts 2 and 3 is made by the viscosity of the silicon oil inside the joint, working on discs 5 and 6; the friction torque depends on the relative speed between 2 and 3.

Newton's law, applied to this case, says that:

$$\tau = \frac{\mu \Delta V}{d} \ , \tag{13.22}$$

where τ is the shear on the surface element dS of two facing discs, d is the gap dimension between discs, ΔV is the local difference in speed and μ is the coefficient of dynamic viscosity; the force acting on the surface element is:

$$\tau r^2 d\alpha \ dr \ , \tag{13.23}$$

where r is the local radius and $rd\alpha$ and dr are the circumferential and radial dimension of the surface element.

If we call r_i and r_e the inside and outside radii of the facing surfaces of two neighboring discs, the total braking torque of the joint bM, at a speed difference $\Delta\Omega$ will be:

$$bM = n \frac{\pi \Delta\Omega \nu \rho}{2d} r_e^4 (1 - \frac{r_i}{r_e})^4 \ , \tag{13.24}$$

where n is the number of facing surfaces of the discs.

The cinematic viscosity coefficient ν is generally used in place of the dynamic coefficient. The relationship between the two coefficients is:

$$\nu = \frac{\mu}{\rho} \ .$$

The value of obtainable viscosity, depending on the quantity of emulsified air in the oil, can range between 30,000 and 100,000 cSt (mm^2/s).

In the Ferguson differential the blocking coefficient therefore depends upon the relative differential speed.

As we have said, the joint can be integrated either in the central or in the axle differential or to be used alone on the propeller shaft.

In the first case the differential could be a spur gear epicycloidal differential with simple or double satellites. If the input shaft is connected to the annulus gear, one half axle will be connected to the sun gear and the other to the carrier; the Ferguson joint is sensitive to the difference in speed between sun gear and carrier.

13.5 DIFFERENTIAL EFFECT ON VEHICLE DYNAMICS

A real differential or a self-locking differential have a beneficial effect on vehicle mobility, because they allow a wheel or an axle on a high traction coefficient ground to receive a traction force greater than that acting on the coupled wheel or axle on low traction coefficient ground.

This advantage has a cost in energy due to the work of friction forces on rotating elements. It also implies variations in traction forces as compared with the ideal reference values. We will see in the following paragraphs the effect of these variations on the vehicle dynamic behavior.

13.5.1 Driving axle differential

Let us first consider the case or rear wheel drive.

We assume that the car is driving on a large steering-pad 200 m in radius at different steady speeds.

The working conditions of the car, with particular reference to driving wheel speed and torque, were calculated by means of a multibody mathematical model. The calculation parameter is the centrifugal acceleration a_y, measured in grams. The overall vehicle mass is about 1,300 kg.

We begin with the case of an ideal differential. In Figs. 13.15 and 13.16 the diagrams of driving wheels torque M and the speed difference between driving wheel $\Delta\Omega$ are represented as a function of the lateral acceleration. The results are as follows:

- If mechanical efficiency is 100%, the torque on the driving wheels is the same, by definition. Torque increases quickly with speed (0.7 g of lateral acceleration corresponds to about 130 km/h of speed on a 200 m steering-pad) because of the increase in driving resistance and the longitudinal components of side slip forces.

- Let us remember that vertical load transfer on driving wheels increases roughly with the square of the speed, being proportional to the acceler-

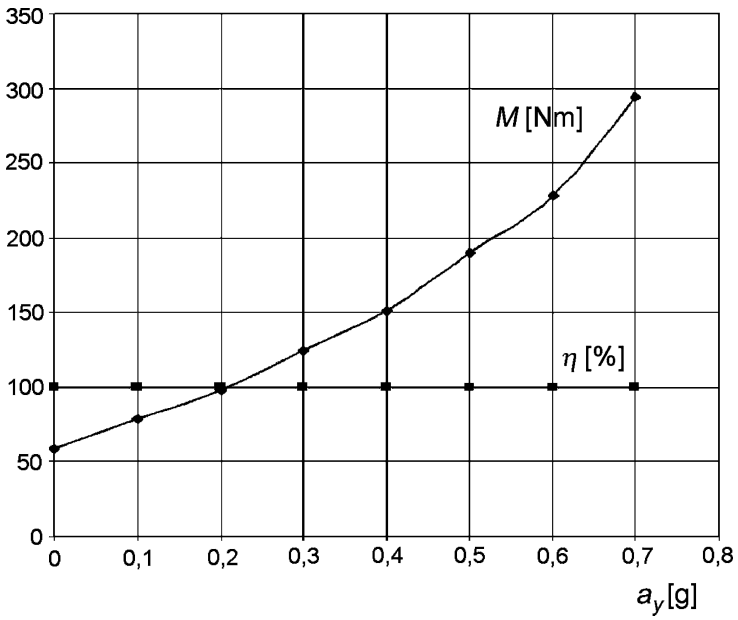


FIGURE 13.15. Diagram of driving wheel torque M, for a rear wheel driven car, on a 200 m radius steering-pad, as a function of centrifugal acceleration a_y, assuming an ideal differential mechanical efficiency η of 100%.

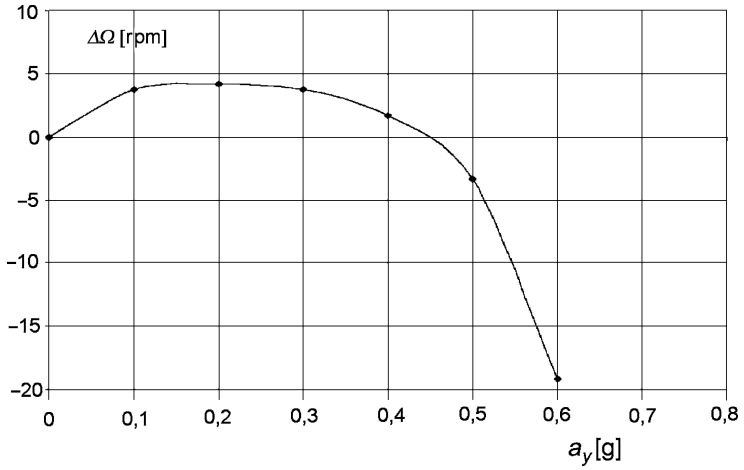


FIGURE 13.16. Diagram of the speed difference $\Delta\Omega$ between the inside driving wheel and the outside driving wheel in the case of the previous figure.

ation a_y. The speed difference $\Delta\Omega$ first increases because of the different wheel path, than decreases because of the fact that the inside wheel, initially slower, experiences increased longitudinal slip, because of the vertical load loss at the same traction force; the speed difference is negative and infinite when the wheel loses its grip.

Assuming that the differential has a constant efficiency smaller than one, the angular speed difference will also allow a torque difference.

The inside wheel, initially slower, will receive a larger torque until the speed difference becomes zero; the torque difference is proportional to total driving torque. Around the point where the speed difference is zero, the driving wheel torque diagrams will cross, leading to the inside wheel receiving a smaller torque, because its speed is faster.

The efficiency diagram will show a cusp with a value of 100% (see Fig. 13.17), where the speed difference goes to zero. The speed difference is zero as long as the differential is locked, because the torque difference is so small.

Let us assume that on the two planet wheels of the differential there is a constant friction force (pre-load). The torque diagrams will cross where longitudinal slip assumes such a value that the speed difference becomes zero.

The efficiency diagram shows the above mentioned cusp and is shaped like a hyperbole, which takes into account the constant value of the friction losses compared to the increasing value of the total torque.

The Torsen and the ZF differential may be simulated, with a good approximation, as constant efficiency differentials. The shape of the curve is similar to the previous and is not shown.

In the case of a viscous differential, shown in Fig. 13.18, the difference between driving wheel torques is not relevant between the origin of the diagram and

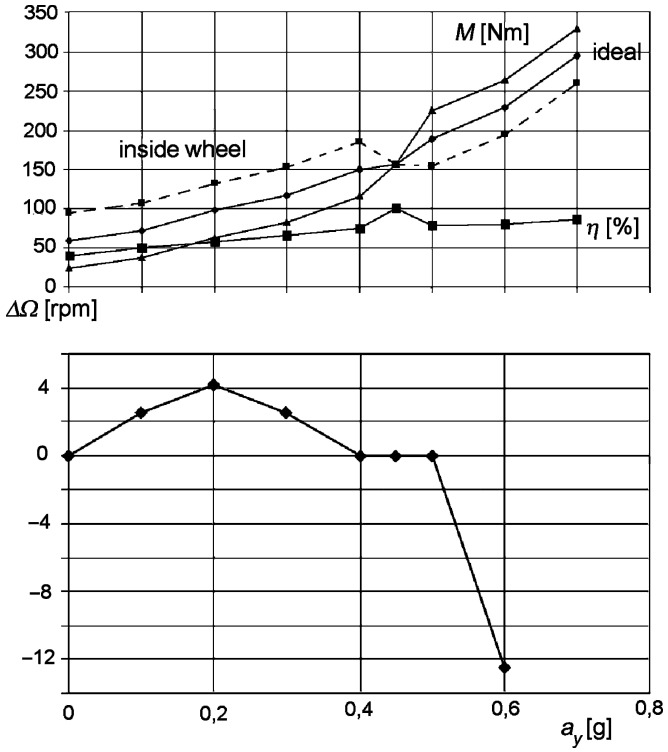


FIGURE 13.17. Upper figure: driving wheel torque M diagram and efficiency η diagram in the case of a pre-loaded differential; torque on inside (dotted line) and outside wheel can be compared with the ideal differential value without mechanical losses. Lower figure: speed difference $\Delta\Omega$ between wheels.

the point where the speed difference becomes zero, because the locking torque is highly dependent on the value of the speed difference.

After the zero point the speed difference increases and therefore the difference in torque increases until it reaches the inside wheel spin point.

Neglecting mechanical losses, efficiency diagram shows two cusps with a value of 100%, at zero acceleration and at the point where the speed difference is zero.

Let us now study the effect of these phenomena on the understeer curve of this car; we will assume as the understeer parameter the difference between the actual steering wheel angle and the steering wheel angle at zero speed (roughly corresponding to the Ackermann steering wheel angle). We limit the study to the case of the ideal differential and of Torsen differential. The understeer diagram is in Fig. 13.19.

We can outline the following facts:

- The zero speed steering angle (the value is shown in the caption of the figure) increases because of the differential friction torque. As a matter of

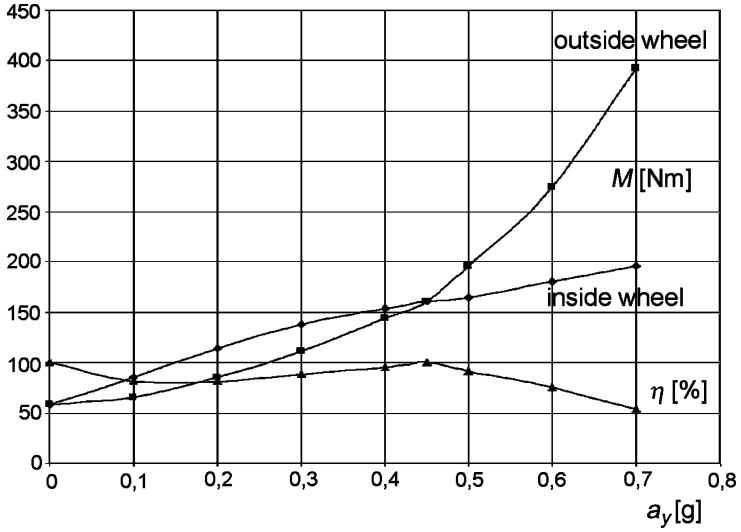


FIGURE 13.18. Diagram of the wheel torque M and the efficiency η of a differential of a rear driven car with Ferguson joint.

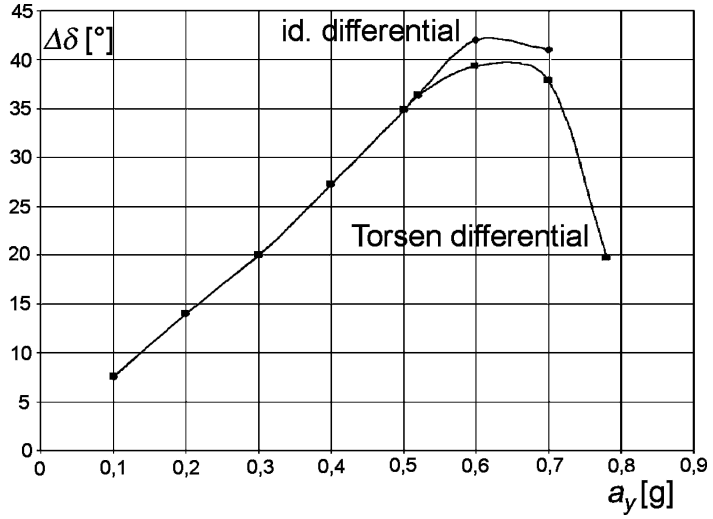


FIGURE 13.19. Understeer index diagram (the understeer index is assumed to be the difference $\Delta\delta$ between the actual steering wheel angle and the zero speed steering angle) of a rear wheel driven car on a 200 m radius steering-pad in the case of ideal differential and Torsen differential. The zero speed steering angle is 13.6 deg for the ideal differential and 14.9 deg for the Torsen differential.

fact, the non-symmetric traction torque induces a yaw moment (understeering direction) at a positive speed difference; the driver must increase the steering wheel side slip angle, in order to equilibrate this torque.

- This additional yaw torque goes to zero where the speed difference becomes zero and becomes, after this point, an oversteering yaw torque.
- The acceleration limit where the rear wheel side slip angle becomes excessive and the steering angle goes to zero (power oversteer) shifts toward higher values because of the lower traction force on the inside wheel.

For small curvature radii steering-pads the diagrams are not dissimilar. The zero speed difference point moves to higher accelerations because the speed difference contribution of the curvature of the path predominates that of the longitudinal slip. The torque difference induced by the differential locking torque is more significant because of the higher longitudinal slips. The steering wheel will become much heavier at low speed; from this standpoint the constant locking coefficient differential is more penalized.

In the case of front wheel driven cars, these calculations can be repeated.

The additional yaw torque (driving torque imbalance) is now applied to the front wheels, with a more significant effect on the steering torque; in addition, the torque imbalance induces an additional steering angle because of the suspension and steering mechanism compliance.

Therefore the application of self-locking differential to front wheel driven cars is seldom seen, with the exception of Ferguson joints with modest self-locking torque.

We can conclude that on a rear wheel driven vehicle (the case of the rear axle of an all wheel driven vehicle may be included) the application of a self-locking differential will increase understeer when entering the curve, but with an ampler exploitation of the road grip.

Poor handling is more evident on small curves or when accelerating the vehicle.

An ideal differential locked by means of a controlled clutch should satisfy the following conditions:

- Simulate the behavior of a no friction differential for high curvature curves at low speed, or when the rolling speed difference is only due to pressure, or when a high grip is not required
- Simulate the behavior of a locked differential during high acceleration turns or on low grip roads

We should not forget that the interconnection torque affects braking with an ABS system. As a matter of fact the prediction of the traction coefficient through the interpretation of the speed difference is affected by the effects on the locking torque; a third requirement of the above differential should be, therefore, to act as a free differential during braking.

13.5.2 Transfer box differential

An all wheel drive transmission system allows higher longitudinal acceleration at a given traction coefficient value.

Let us recall the equilibrium equations of a longitudinally accelerating vehicle that will be better explained in the second volume; we call:

h	the height on the ground of the center of gravity,
l	the wheelbase,
F_{z1}, F_{z2}	the vertical loads on front and rear axles,
μ_x	the peak value of the traction coefficient,
P	the vehicle weight,
a_x	the longitudinal acceleration,
F_{x1}, F_{x2}	the traction forces at the adhesion limit on the front and rear axles.

On a front wheel driven vehicle there will be a vertical load transfer due to the longitudinal acceleration:

$$\Delta F_z = a_x \left(\frac{Ph}{lg} \right) , \tag{13.25}$$

therefore:

$$F_{x1} = \mu_x (F_{z1} - \Delta F_z) , \tag{13.26}$$

$$a_x = g \frac{F_{x1}}{P} = \mu_x g \frac{F_{z1}}{P\,(1 + h\mu_x / l)} ; \tag{13.27}$$

in the same way on a rear wheel driven vehicle, it will be:

$$a_x = g \left(1 - \frac{F_{z1}}{P} \right) \frac{\mu_x}{1 - \mu_x \frac{h}{l}} . \tag{13.28}$$

In the case of an all wheel driven vehicle, we introduce the parameter i, identifying the characteristics of the transfer differential:

$$i = \frac{F_{x1}}{F_{x1} + F_{x2}} . \tag{13.29}$$

The former formulae become:

$$a_x = g \frac{F_{x1}}{P} \frac{\mu_x}{i + \mu_x \frac{h}{l}} , \text{ or} \tag{13.30}$$

$$a_x = g \left(1 - \frac{F_{x1}}{P} \right) \frac{\mu_x}{1 - i - \mu_x \frac{h}{l}} . \tag{13.31}$$

The choice between the first and the second equation depends on whether the front or rear axle spins first.

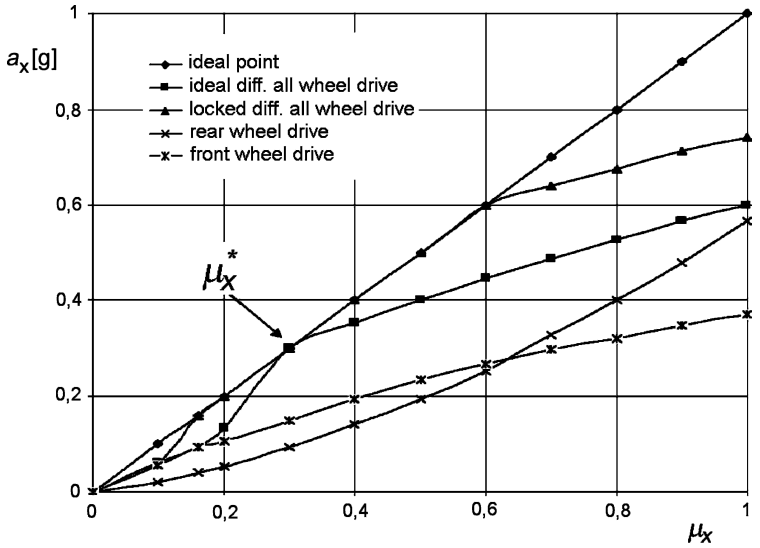


FIGURE 13.20. Diagram of the maximum longitudinal acceleration a_x (on the vertical axis, measured in grams), as a function of the longitudinal traction coefficient μ_x (on the horizontal axis), in the case of an ideal material point, a front wheel drive, a rear wheel drive and an all wheel drive with no friction differential and for a locked differential with constant locking coefficient. The higher the locking coefficient, the wider the extension where the ideal material point behavior is copied.

Figure 13.20 compares the diagram of a_x as a function of μ_x in the three considered cases; the advantage of the all wheel drive is evident, where, in any case, the ideal condition of:

$$a_x = g\mu_x,$$

is reached at one point only.

The value of the traction coefficient μ_x^* where the traction force is the same as in a material point with no load transfer can be derived when the two previous equations are equal:

$$\mu_x^* = \frac{l}{h}\left(\frac{F_{x1}}{P} - i\right) \ . \tag{13.32}$$

For acceleration lower than the $g\mu_x^*$ value, the rear wheels will spin, because the load transfer is not sufficient to guarantee the necessary traction force; the opposite occurs for higher accelerations.

If the transfer differential can be modelled as a device with a locking coefficient b, we can substitute in the previous formulae a fictitious i' coefficient equal to:

$$i' = i - \frac{b}{2}$$

and i'' equal to:

$$i'' = i + \frac{b}{2},$$

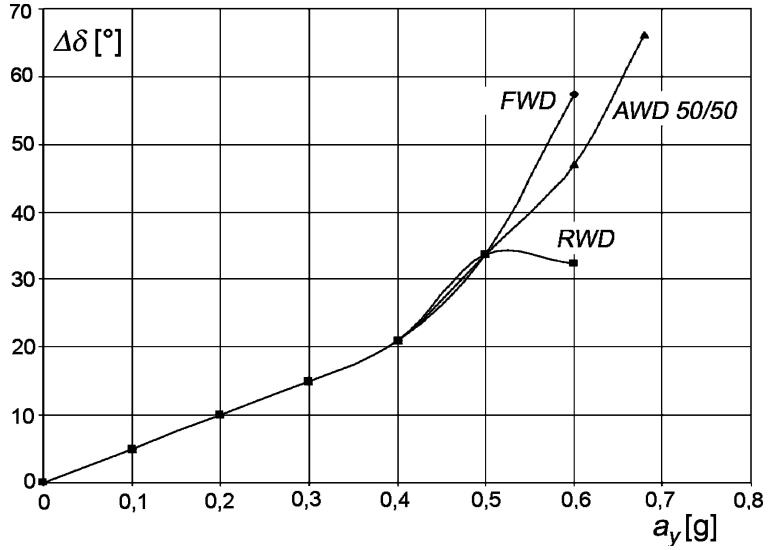


FIGURE 13.21. Diagram of the understeer index $\Delta\delta$ as a function of the lateral acceleration a_y for the car in the previous examples, with a front wheel FWD, rear wheel RWD and all wheel drive AWD (50% of torque on each axle) with no friction transfer differential.

to take into account the driving torque increase or decrease on the axles caused by the locking torque of the transfer differential; in the diagram representing the maximum acceleration as a function of the longitudinal traction coefficient, the curve will match the ideal straight line for some extension; the larger the value of b, the wider the extension.

Let us examine finally the effect of all wheel drive on the understeer coefficient, again using the steering-pad simulation.

An ideal transfer differential splitting the traction torque into equal parts is considered; in Fig. 13.21 the understeer curves relative to this case are compared with those of the front wheel and rear wheel drive.

On rear wheel drives, at high acceleration when the traction force is relevant, the response tends to oversteer because of the high value of the rear side slip angles; on front wheel drives, the high value of the side slip angles raises the natural vehicle understeer to unacceptable values (the maximum acceleration is limited by the steering rack allowed stroke).

The acceleration value where the vehicle response degenerates is almost the same for front and rear wheel driven cars. All wheel drive reduces side slip angle values and is characterized by a better response. In the reported case, an acceptable understeering behavior is also present at high values of acceleration, with increased manoeuvrability for the car.

14
SHAFTS AND JOINTS

Transmission shafts are used to apply torque to those driven components whose rotation axis cannot be perfectly aligned with their driving counterpart.

The following connections fit this description:

- Gearbox output shaft or reducer output shaft with final drive, for conventional drive vehicles using either rigid axle or independent suspensions; this transmission shaft is also called the *propeller shaft*.
- Gearbox output shaft with reducer or transfer box input shaft, when they are not rigidly connected because they are mounted separately on the chassis structure.
- Engine shaft and gearbox input shaft, when they are separated.
- Differential output shaft with driving wheels in all independent suspensions; these are also called *half shafts*.

It is evident in such cases that transmission shafts must be capable of working with noticeable offsets and angles, because of the suspension stroke and, for front wheels, because of steering angles.

Nor can transmission components directly connected to the chassis structure, as, for example, gearbox and differential box in conventional drive vehicles with independent suspensions, be joined by rigid flanges, because of the flexibility of the chassis and dedicated elastic suspensions, addressed to vibration dampening.

Shafts fall under two different categories: Propeller shaft and half shafts.

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

14.1 PROPELLER SHAFTS

A typical propeller shaft can be exemplified by an industrial vehicle application, as shown in Fig. 14.1.

The final drive on the rigid axle is connected with the gearbox output shaft through a two piece propeller shaft: The swinging shaft 1 and the fixed shaft 2. An intermediate bearing 3 holds the fixed shaft aligned with the gearbox output shaft and the connection with the swinging shaft.

The swinging shaft is subject to changing its inclination angle with reference to the side beams of the framework because of rear suspension bounce; to evaluate the offset to be covered, lateral displacement due to suspension and chassis flexibility should be considered.

Because the reaction to the driving or braking torque applied to the axle is also absorbed by the suspension elastic elements, we can expect axle rotation around the y axis; see Part I above the S deformation of leaf springs.

The two piece design is demanded by the relevant distance between the flanges to be connected, considering the natural bending frequencies of the shaft; it is desirable that these frequencies be quite dissimilar from the natural bending frequencies of the chassis to avoid annoying booms.

For this purpose an elastic suspension is also provided for the intermediate bearing.

For shorter vehicles, such as semi-trailer tractors or light duty industrial vehicles, a single piece propeller shaft can be used.

The natural torsion frequency of the system, including power train, transmission shaft and vehicle has a fundamental importance on vehicle comfort and driveability; this natural frequency is directly influenced by the propeller shaft.

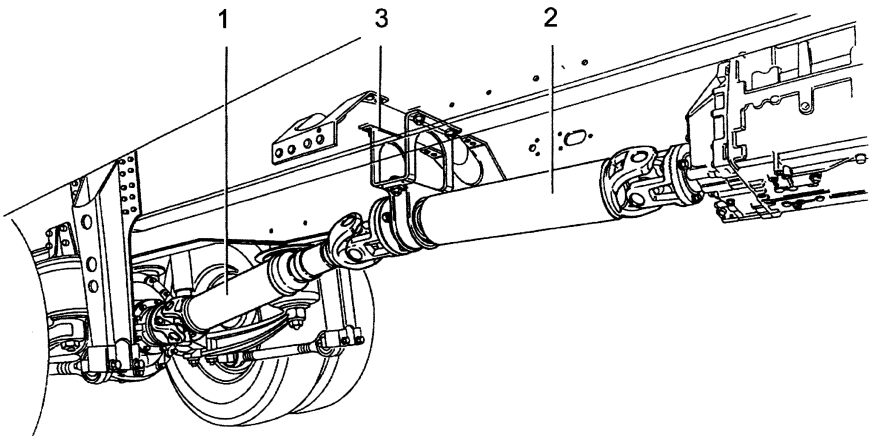


FIGURE 14.1. Two piece propeller shaft for the rigid rear axle of a heavy duty industrial vehicle (Iveco).

A simple model of the system could include two flywheels of inertia J_m and J_v representing respectively the power train inertia (engine and gearbox, at a given speed) and the inertia of the vehicle; these flywheels are connected through a transmission shaft, whose length is L.

If G is the shear module of the shaft material, and I_p the polar inertia of its cross section, we will have, at resonance condition:

$$\omega = \sqrt{\frac{GI_p}{J_m L} + \frac{GI_p}{J_v L}} \,, \tag{14.1}$$

where ω[1] is the natural torsional frequency of the transmission shaft.

As we know, internal combustion engines deliver a torque that changes over time, because of the shape of the pressure cycle and because this pressure is transferred to the shaft through a crank mechanism; also the engine rotating and reciprocating masses of the engine introduce a further cause for driving torque variation over time.

If we break down the resulting torque into harmonics, it will be possible to identify engine speeds introducing resonance between the driving torque on the gearbox output shaft and the natural frequencies of the transmission shaft; the desired result is that this resonance condition is beyond the vehicle maximum speed or occurs at speeds so low that they do not affect the conditions of vehicle use in a practical way.

We should remember that other torsional excitations come from the joints connecting the different pieces of the propeller shaft.

If we now model the propeller shaft with a constant section beam with section S and inertia moment I_f, and we designate as E the elasticity modulus and as ρ the density of the shaft material, we can also calculate:

$$\omega_c = \frac{\pi^2}{L^2} \sqrt[2]{\frac{EI_f}{\rho S}} \,, \tag{14.2}$$

to derive the natural bending frequency of the propeller shaft on its bearings.[2]

This approach is oversimplified, because it does not take into account the contribution of other elastic elements of the transmission system, such as clutch torsional dampers, tires and suspensions, for torsion or bearing, and power train suspensions for bending; a more complete model will be introduced in Part V of the second volume.

In any case, the issue of natural frequencies suggests sizing transmission shafts with rigid cross section and very light design; tubular sections are, in consequence, the best choice.

---

[1] In this book vibration frequency is indicated by ω, while Ω is used for angular speeds. Frequencies will always be measured in coherent units (rad/s) and no distinction will be made between frequency and pulsation.

[2] A more complete approach to the issue of natural frequencies and critical speeds can be found, for example, in the book G. Genta, *Dynamics of Rotating Systems*, Springer, New York, 2005.

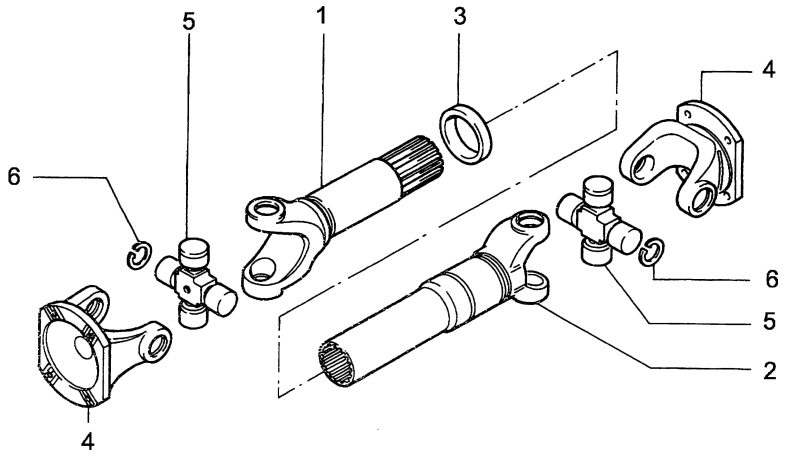


FIGURE 14.2. Exploded view of the moving part of a propeller shaft with universal Hooke joints.

High resistance steel or aluminum are used; sometime on sports cars, filament wound composite eposidic plastic materials with kevlar fibers are also applied.

The propeller shaft of conventional drive cars with suspended differential is designed with the same criteria.

With these applications a universal Hooke joint is applied, as shown in Fig. 14.2.

The moving piece of the propeller shaft is made by two tubes 1 and 2, connected by a spline that allows length variations due to the motion of the axle suspension. The spline is lubricated by a grease sealed by a rubber ring 3. The two ends of the moving piece show the *yokes* of the universal joints.

Two other yokes 4 are flanged to the gearbox output shaft and to the differential input shaft or fixed transmission piece, if any.

Two *trunnion crosses* 5 connect the yokes; the connection between crosses and yokes is usually made by sealed needle bearings, kept in place by Seeger rings 6. We comment in a following paragraph on the operation of the universal joint.

14.2 HALF SHAFTS

The half shaft layout of an independent wheel suspension for a front wheel driven car is shown in Fig. 14.3.

AS the driving axle is also a steering axle, joints must be designed for large working angles. The transversely mounted power train always implies half shafts of different length.

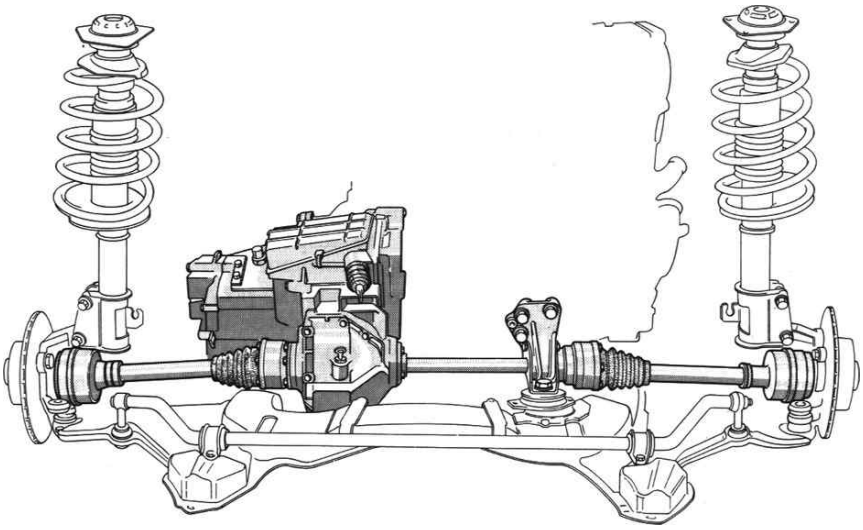


FIGURE 14.3. Complete transmission for a mid-size front wheel driven car with transversal engine; the transmission is seen from the back (Lancia).

These different lengths can be a problem, if because of them the half shafts feature a different torsional stiffness; this occurs in wide cars with powerful engines. In fact, under high values of driving torque, self-steering moments could be applied to the steering mechanism, because the stiffer half shaft transmits higher torque. The problem can be solved by using bigger cross sections on the longer half shaft.

By this design choice, the different geometry of the two half shafts could cause the natural bending frequency of the longer half shaft to be too low; to avoid this inconvenience, as with propeller shafts, the longer half shaft is divided into two pieces, the first fixed to the power train assembly.

What was written in the previous paragraph about torsional and bending natural frequencies could be repeated in this case, but the application of tubular sections here may prove impossible because of the limited space.

It can therefore happen that a half shaft natural frequency couples with a tire vertical natural frequency, with resonance. This problem can be solved by applying additional mass to the half shaft; this can be accomplished by an iron ring press fit on a rubber bushing with suitable elasticity. This systems works as a damper tuned on its natural frequency.

The study of torsional and bending vibrations must take into account the entire power train suspension, now affected also by reaction to the driving torque on the front axle; we approach this study in Part V.

Because of the limited space and wide angles, constant velocity joints based upon the Rzeppa principle are applied to half shafts rather than universal joints.

In independent rear suspensions technical solutions are similar to these of front driving suspensions.

14.3 UNIVERSAL JOINTS

The design of a universal joint has been already described in the previous Fig. 14.2; it is easy to understand why a universal joint alone is not a constant speed[3] joint.

Let us consider the scheme in the upper part of Fig. 14.4 and in particular the left joint.

If θ is the rotation angle of the first shaft piece 1, before the joint, at the time t, φ is the corresponding rotation angle of the second shaft piece 2, after the joint and α is the angle between the center line of the shafts 1 and 2, we can write that:

$$\tan(\varphi) = \frac{\tan(\theta)}{\cos(\alpha)} \,. \tag{14.3}$$

If we derive the two members of the above equation, we obtain:

$$\frac{d\varphi}{dt} = \frac{\cos(\alpha)}{1 - \sin^2(\alpha)\cos^2(\theta)} \frac{d\theta}{dt} \,. \tag{14.4}$$

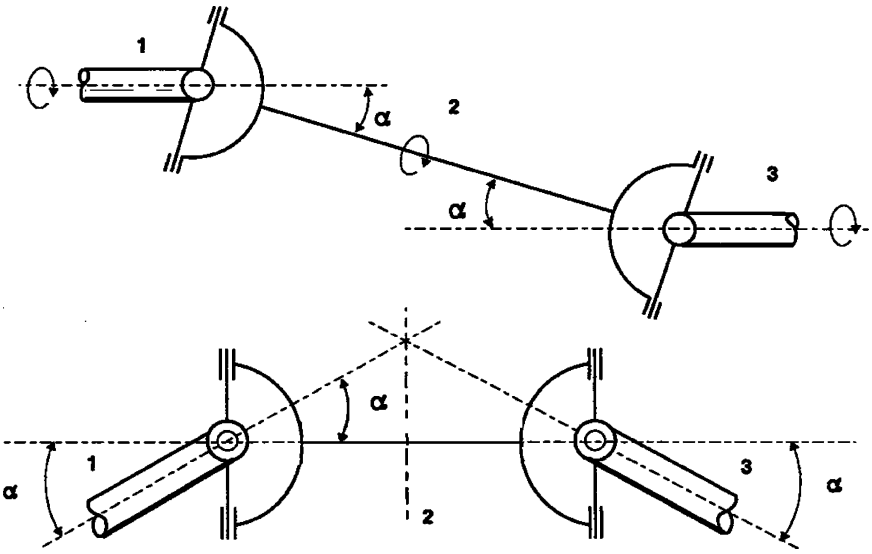


FIGURE 14.4. Universal joint layout necessary to obtain a constant speed transmission; the inertia moments of the the shaft mass between the joints will generate, in any case, a periodic output torque.

[3]In a constant speed joint, the ratio between input and output speeds is always independent of angular position; in particular, if the input speed is constant, the output speed will also be constant and equal.

Therefore if the input speed

$$\frac{d\theta}{dt}$$

is constant, the output speed

$$\frac{d\varphi}{dt}$$

will not be and there will be a periodic function of θ expressing the transmission ratio between input and output shafts; this function will depend on α. Constant speed conditions are obtained only for $\alpha = 0$.

But if we install on the second shaft piece a second universal joint, connecting the third shaft piece 3, we can, for the same equation, obtain a constant speed transmission if the second joint has the same working angle α and its input yoke is on the same plane as the output yoke of the first joint. This task can be accomplished in two ways, as in the upper and lower part of the figure.

This sentence is only true in the ideal case where the rotating mass of the second shaft piece 2 is negligible in comparison with the total rotary mass of the system; if not the inconstant speed of the shaft 2 will introduce torque variations in the system that will not conserve its constant velocity.

The configuration applied to propeller shafts is that shown in the upper part of the figure and implies that the shafts 1 and 3 are always parallel; this condition must be obtained by a suitable elasto-kinematic behavior of the rear axle suspension. This condition is only approximately obtained by a leaf spring suspension because of the S deformation of springs owing to the driving torque.

The second layout in the lower part of the figure is used particularly by the half shafts of front driving axles of some off-road vehicles, featuring rigid axle suspension, where the high value of the torque does not allow constant speed joints to be used.

Figure 14.5 shows an application of this kind. The stub axle and the axle are articulated through the king-pin for steering; the king-pin axis must be coincident with the middle of the shaft 2; this is very short, reduced to the minimum necessary to contain the needle bearing of the trunnion crosses.

In this arrangement, the steering motion induces two equal working angles in the two universal joints, as demanded by the previous scheme. The axle and the stub have a particular shape building up a sealed rotula to protect the transmission.

Sizing universal joints is not difficult; usually they are selected in a catalog, taking into account transmission torque and working angles. It is important to consider that torque is limited by working angles and there are also geometric limits, due to interference between yokes and cross that must be verified before the application.

It also important to avoid applying transmission shafts with working angles that are too small; in fact, small working angles can locally wear needle bearings because of limited rotation.

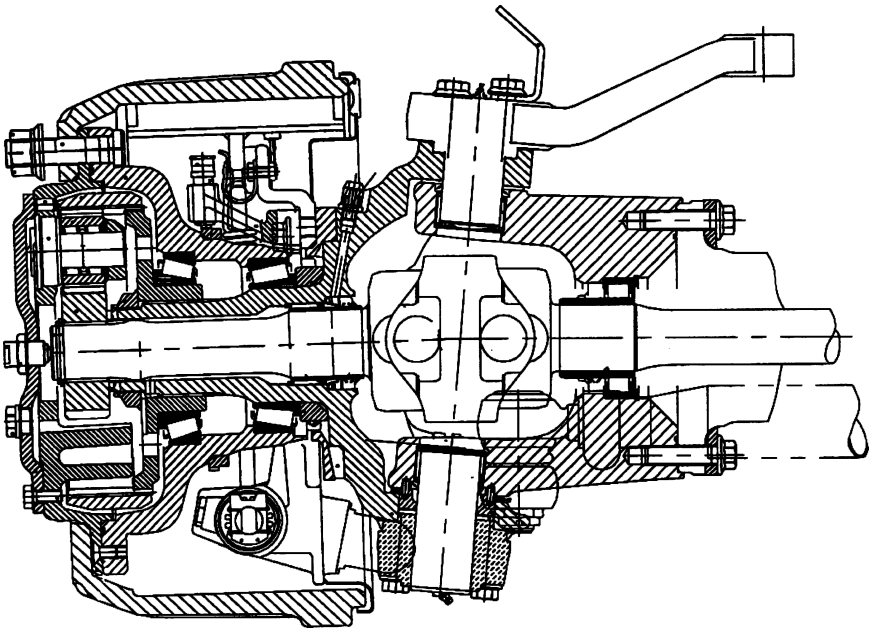


FIGURE 14.5. Scheme for a universal joint transmission for the front rigid steering axle of an off-road industrial vehicle; the two yokes of the shaft 2 are almost coincident (Iveco).

14.4 CONSTANT SPEED JOINTS

The double universal joint cannot be used in the transmission of an independent suspension to the steering wheels; this solution has been applied in the past for non-steering rear wheels: This solution does not appear acceptable today because of the limitations on camber recovery, in order to keep the first and third shafts of the transmission parallel.

In independent wheel suspensions only constant velocity joints of the Rzeppa type are applied, such as those shown in the upper part of Fig. 14.6; the scheme represents a section of a complete half shaft for a front wheel driven car. The wheel, not represented, is on the right of this scheme.

The joint on the wheel side must be able to rotate because of two components: The steering angle, in an almost horizontal plane, and the angle imposed by the suspension motion (stroke and camber) in an almost vertical plane; the total working angle allowed by these joints is about 45 deg.

The other joint on the gearbox side is subject only to the angles caused by the suspension stroke in an almost vertical plane and by the steering angle in an almost horizontal plane; in fact, because the joint on the wheel side cannot be set on the king-pin axis, the wheel steering motion will move the joint center on an almost horizontal plane; the total working angle allowed for these joints is about 20 deg.

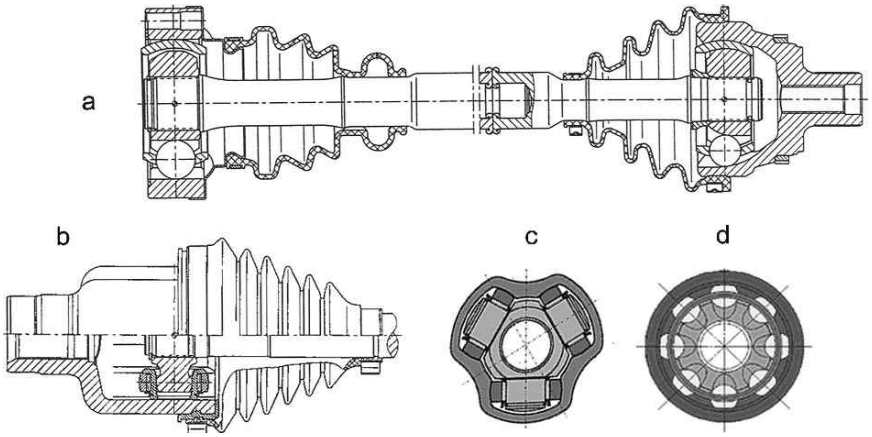


FIGURE 14.6. Complete half shaft for a front wheel driven car (a) with a fixed joint on the right and a sliding joint on the left. Scheme (b) represents a tripod joint that can be used in place of a sliding Rzeppa joint. Details (c) and (d) show the cross sections of these joints.

The wheel steering and the suspension stroke impose on the joint on the wheel side a trajectory not coincident with a circle centered on the joint on the gearbox side; for this reason a sliding joint must be applied that will allow up to 50 mm of displacement.

Rzeppa sliding and fixed joints are made with a crown of balls engaged in corresponding grooves on the inside nut and on the outside cup, where the wheel or the differential output shafts are connected. The balls are kept on the same plane by a cage, made by a ring with holes.

Scheme d in the figure shows a section of the fixed joint, not dissimilar in this view from the sliding joint section.

The shape of the ball grooves on the nut and cup is such as to determine four different contact points where the surface in contact is almost perpendicular to the force to be exchanged.

The intersection of the plane containing the ball centers with the joint rotation axis determines the joint center or the half shaft articulation point; the cup of the joint on the wheel side uses a groove curvature that avoids ball displacements along the rotation axis, while the cup on the gearbox side has straight grooves to allow the half shaft freedom to slide.

It is important to verify that no suspension position (usually the most dangerous one is at rebound with full steering angle) can result in a ball exiting its groove.

Cups are sealed with rubber bellows, rotating with the half shaft, that keep the balls lubricated for life.

The constant speed behavior of these joints can be explained by Fig. 14.7. In this scheme only one ball 3 is represented, engaged with two grooves at the same time; these grooves are fixed respectively to the two parts of the joint 1

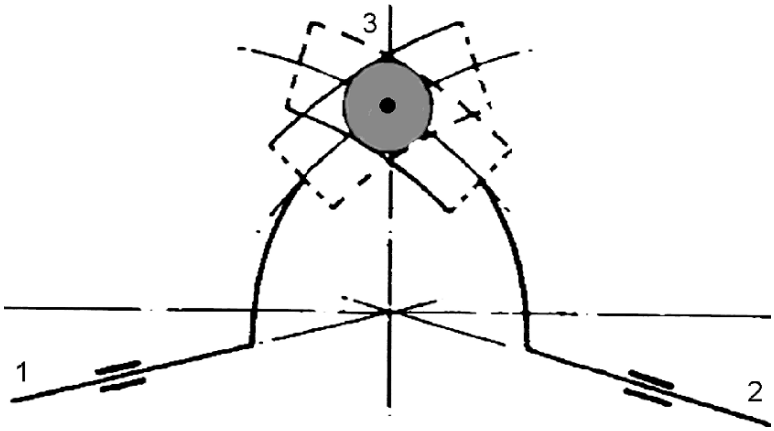


FIGURE 14.7. Scheme of a constant speed Rzeppa joint; only one ball 3 is represented, engaged with the two grooves on shafts 1 and 2.

and 2; only the part of these grooves represented with a full line actually exists on the nut and cup.

The shape of the groove is such as to cause any ball to have its center on the bisector plane of the angle made by the two rotation axes of the shafts of the joint; in this condition the distance between each ball and the two rotation axes will be constant, determining the identical rotation speed of the two axes. During joint rotation the balls will be subject to displacements along their grooves proportional to the joint working angle.

From the description of their operation it appears that friction of the ball joints must be carefully considered; in fact, in a complete revolution of the joint each ball is displaced along its groove in proportion to the working angle; this displacement cannot described by pure rolling of the balls in their grooves.

The friction causes a small energy loss in the transmission and an exchange of forces in the transversal direction, because of suspension and power train motions; these forces can cause vibrations and discomfort that is not efficiently filtered by the powertrain suspension.

For this reason the limit of the force necessary to have the joint sliding under torque must be carefully specified and taken under control; the tripod joint is an attempt to solve this problem with a different architecture.

A scheme of this joint is represented in the scheme b in Fig. 14.6. The balls are replaced by three rollers; these are mounted on pins with needle bearings and can roll in the straight grooves of the cup. This design is suitable to contain friction caused by joint sliding motion under torque.

15
AUTOMATIC GEARBOXES

15.1 GENERAL ISSUES

Unlike their manual counterpart, automatic gearboxes employ a wide variety of technical solutions, a possible demonstration that this kind of gearbox has yet to attain his technical maturity.

Automatic gearboxes were created for comfortable family cars with high displacement engines. They are now increasing their market share in small cars and sports cars, a market where customers expectations are more addressed to performance, economy and drivability than to comfort; in addition, a higher attention is devoted to emissions and energy conservation. Technical solutions to these problems have proliferated, with no final winner yet determined.

The technical solutions described in this chapter are those that have attained the widest diffusion in the automotive market; some solutions specific to industrial vehicles will also be outlined.

An important topic for automatic gearbox study is their control system; in modern cars automatic gearboxes primarily use electronic control systems.

Their study is usually undertaken in a course on electric and electronic systems.

Nevertheless we will examine automatic gearbox control strategies, which can be defined as rules and methods that are adopted in control systems to decide and actuate speed shifts.

Hybrid vehicle automatic transmissions are not considered here; they feature the further function of determining, which is the most convenient energy source in a given situation (prime engine, energy storage system, wheels).

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

15.1.1 Automation level

In the historical perspective, the first goal to be reached by automatic gearboxes has been to improve driver and passenger comfort.

The driver is relieved of the physical effort of operating the clutch pedal during start-up and speed changes and shifting the gears; also the need to identify the most suitable moment for shifting and to coordinate operations, especially in an uphill start, is eliminated.

For passengers the objective has been to limit the *jerk* (derivative of the vehicle acceleration) during gear shifts and starts; the human body is quite sensitive to this parameter making it relevant to travel comfort.

If we consider automatic gearboxes from the standpoint of automation, we can identify the following types:

- *Full automatic* gearboxes; shifting speed and start-up functions are both carried out automatically, with mechanisms developed for this purpose. Among different available working modes, there is usually a semi-automatic mode where speed ratio choice is left to the driver, the shifting sequence being entirely automated.

- *Semi-automatic* gearboxes; one of the above automatic functions is partly or totally missing. For example, start-up device operation is automatic, but not the gearshift; or this second function is automatic, but decisions about the most suitable shifting moment are left to the driver. In this second version, the accelerator pedal must be operated by the driver in coordination with clutch operation, to avoid engine over speed or stall or any kind of shock. This kind of gearbox no longer interests car designer, because it does not represent a favorable trade-off between cost and benefit. It is, however, adopted on some industrial vehicles, because of the effort of performing commands manually; these gearboxes sometimes share most components with their manual counterparts.

- *Robotized* or *automated* gearboxes; mechanisms are deliberately derived from manual gearboxes, including the friction clutch. These cannot be defined as semiautomatic gearboxes, like the previous, because their control systems can handle all automatic functions. The decision to use a modified manual gearbox can be justified by other goals, such as reducing the cost of production and operations. The former is reduced because existing production facilities can be used. The effect on the latter is due to reduction of mechanical losses in comparison with automatic gearboxes. We should not forget, finally, a positive impact on improving sports car performance, because shifting and start-up times can be reduced to figures not readily available even to professional drivers.

This last category of gearboxes is attaining a significant market share and could have further development in coming years.

15.1.2 Gearshift mode

An important aspect conditioning automatic gearbox configuration is the gearshift mode; we usually distinguish between gear shifts with and without power interruption or *powershifts*.

In the first case gears are shifted as in a manual gearbox using synchronized or non-synchronized dog clutches; during gear shift the power flow going through the gearbox must be interrupted to allow the disengagement of the existing speed and the engagement of the next.

The power available to move the vehicle goes to zero during this manoeuvre and the vehicle slows; the jerk can be significant.

The upper part of Fig. 15.1, shows a qualitative diagram of traction force and vehicle speed during a gear shift with power interruption.

The lower part of the same figure shows the case of a powershift gearbox. It is possible to reduce the jerk to a minimum and continue to accelerate the vehicle if the manoeuvre occurs without interruption of power. This result can be achieved on stepped gearboxes if each synchronizer is replaced with a clutch and disengagements and engagements overlap; it can be easily understood that gear shift sequence implies a more significant power dissipation than for a conventional

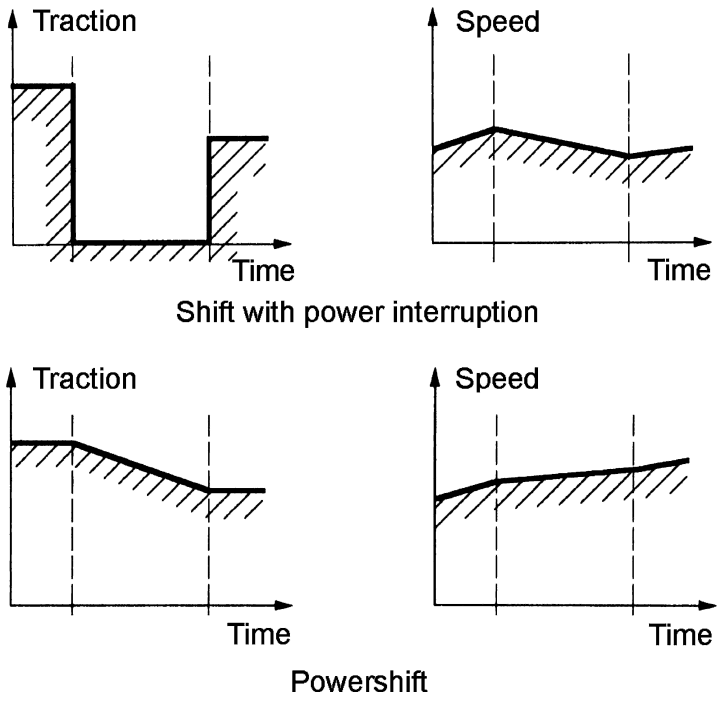


FIGURE 15.1. Qualitative diagrams of traction force and speed as a function of time for gear shift with interruption of power and powershifts.

synchronizer. In this procedure shifting time is not as important and can be compromised in favour of better driving comfort.

We should note that on automatic gearboxes with power interruption, synchronizers may be avoided, if an external power source is used: This can be the engine that can speed up itself and rotating parts to be synchronized with a double clutching manoeuvre, during down shifts. A brake on the gearbox slows the same parts down during up shifts.

This option is not conditioned solely by costs but also by reduction in shifting time; in this case a drive-by-wire accelerator system is mandatory.

On continuously variable gearboxes, shifting manoeuvres are, by definition, of the powershift type, in consideration of the small difference between previous and next transmission ratios; the wasted energy during gear shift can be minimal.

15.1.3 Stepped and continuously variable gearboxes

Automatic gearboxes are called stepped if they have available a limited number of transmission ratios (*steps*) as in manual gearboxes. The total number of speeds is seldom larger than six in consideration of the increased mechanical complexity.

When designing the number of speeds on an automatic gearbox with torque converter, it is useful to remember that this start-up device is itself a continuously variable transmission.

In this case it is usual to avoid the first speed of the corresponding manual gearbox; its function is performed by the torque converter from stall to lock-up and therefore the number of available speeds is reduced by one unit without affecting vehicle performance. A four speed torque converter automatic gearbox can be compared with a five gear manual gearbox with clutch and so on.

Stepless or *continuously variable transmissions*, CVT for short, feature an unlimited number of transmission ratios between a lower and an upper limit.

Mechanisms that perform this function are developed for the purpose and will be explained later. An important design parameter for CVT is the so-called *range*, which is the ratio between the highest and the lowest available transmission ratio. When designing the CVT range, it must be kept in mind that it should be the same as for stepped gearboxes for the same vehicle, taking into account what was said about torque converter application. A CVT with very large range (with very short low ratio) could avoid a start up device.

Gear trains used on stepped automatic gearboxes can be classified according to two categories:

- Gear wheels with fixed rotation axis, similar to those adopted on a manual gearbox; this solution is simple but seldom adopted on powershift gearboxes because of the installation space to be left for clutches.

- Epicycloidal gear trains of different configuration that in association with band brakes or multi disc clutches can achieve different ratios including reverse speed, always featuring coaxial input and output shafts.

15.2 CAR GEARBOXES WITH FIXED ROTATION AXIS

15.2.1 Synchronizer gearboxes

Automatic gearboxes with synchronizers are often derived from manual gearboxes.

In the past, manual gearboxes with torque converters were sometimes applied. In these gearboxes the torque converter was coupled to a clutch and worked as an automatic CVT from stall to lock-up. The clutch, with an electromagnetic actuator, would open automatically at any speed change.

Selection and engagement manoeuvres were manual, performed via the shift stick; the electromagnet of the clutch was switched by a sensor capable of measuring the effort applied to the shift stick knob.

This kind of semi-automatic gearbox eliminated the clutch pedal only, the shift stick remaining manual. These gearboxes no longer exist on cars but could be reconsidered for future use.

Automatic gearboxes, whose mechanisms are shared with those of manual gearboxes with friction clutch, are now present on the market in growing volume; their shifting mechanisms are, on the contrary, quite specific.

External shifting mechanisms are electrically or electrohydraulically servo actuated. Those of the first kind look less appropriate to a quick response. As a matter of fact, hydraulic systems can feature higher power peaks, through the use of pressure accumulators; electric 12 V actuation is, on the contrary, limited by battery capacity and the dimensions of wiring harness.

On hydraulic systems, the external shifting mechanism is made out of two categories of components: A group providing for hydraulic energy generation, regulation and distribution, and a set of actuators. Actuators are as numerous as the displacements to be made: One for the clutch and two, at least, for the gearbox (one for selection and one for engagement).

Figure 15.2 presents a system of this kind; a hydraulic actuator 1 engages and disengages the clutch, while a second double actuator provides for shifting and turning the shaft 2, for engagement and selection motions; the hydraulic group 3 includes an oil pump, a pressure accumulator and electric actuation valves. The selection mechanism is handled by a rack that can raise the shaft 2 matching the finger with the three selection gates (first and second speed, third and fourth speed, fifth and reverse speed).

The second actuator can rotate the same shaft to obtain engagement and disengagement. It must feature three different positions: The engagement of the two neighboring speeds and the intermediate idle position.

In the next Fig. 15.3 actuators are illustrated in greater detail. The clutch actuator 1 is moved by pressure acting on its back chamber (coil spring side): oil pressure operates disengagement, while engagement is obtained by opening a controlled exhaust; the coil provides for actuator return.

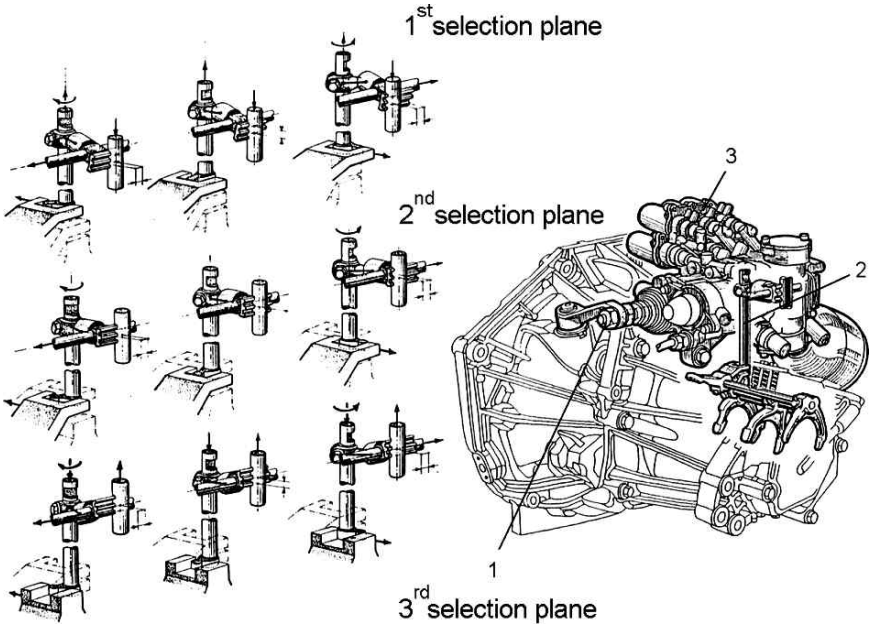


FIGURE 15.2. Scheme of an electro-hydraulically actuated internal shifting mechanism; a hydraulic actuator 1 disengages and engages the clutch, while a second combined actuator provides for rotation and shift of the shaft 2; a hydraulic group 3 includes an oil pump, a pressure accumulator and electric actuation valves (Marelli).

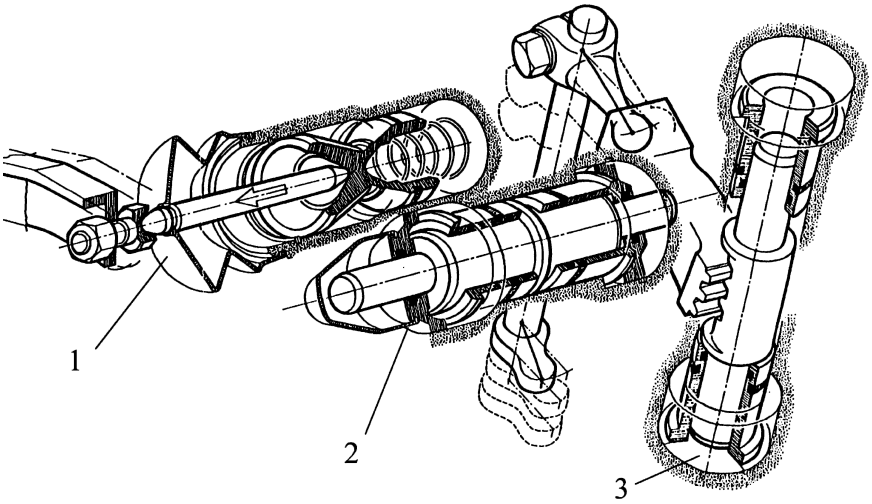


FIGURE 15.3. Hydraulic actuation group for a gearbox; 1 is the clutch actuator; 2 and 3 the engagement and selection actuators (Marelli).

Pressure control must be performed through a proportional valve, suitable for controlling actuation and exhaust pressure to obtain a suitable actuation speed (i.e.: slow start-up, quick start-up, speed change, etc.). Clutch actuators must feature two stable rest positions (disengaged and engaged clutch) made by stroke limiters; the control system must recognize driven plate wear and adjust the incipient engagement point accordingly.

Selection actuator 3 and engagement actuator 2 must feature three different stable rest positions (for three selection planes, for idle and two engaged neighboring speeds). These positions remain almost unchanged during the useful life of the gearbox; they must be adjusted only at the assembly line. Contrary to clutch actuators, engagement and selection actuators have no return coil springs and must therefore receive oil pressure on two piston sides.

Figure 15.4 shows the design feature which allows an intermediate stable piston position to be obtained. Two bushings 2 and 3 are invested on the double effect piston 1, inside the cylinder 4; they are free to move inside the cylinder.

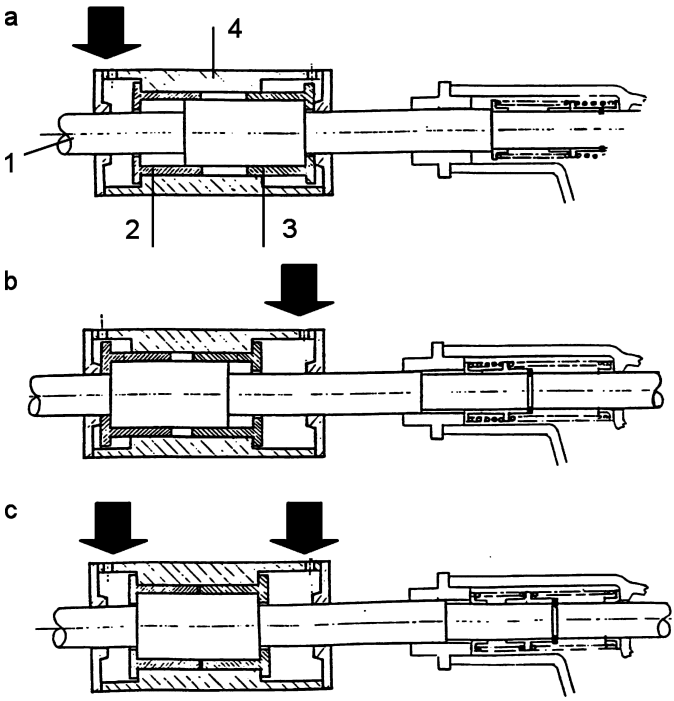


FIGURE 15.4. Actuator piston with three stable positions. Two bushings 2 and 3 are set inside the cylinder; stroke limiters on the bushings determine the intermediate stable position. While feeding the oil to the left chamber (scheme a), a stable position with right full stroke can be obtained; feeding the oil to the right chamber (scheme b), a stable position with left full stroke can be obtained. Feeding both chambers (scheme c) brings the two bushings to the end of their stroke, obtaining an intermediate stable position (Marelli).

Their stroke is limited by a collar that also determines the intermediate stable position (scheme c). Feeding the oil to the left chamber (scheme a) the right stable position can be obtained; feeding the oil to the right chamber (scheme b) the left stable position can be obtained. Feeding the two chambers at the same time (scheme c) brings the two bushings to the end of their stroke, obtaining the intermediate stable position.

In the generator, regulator and distributor group there is a pressure accumulator that regulates pressure at a value almost independent of the actual flow rate required by the actuators. To achieve this, the pump is driven on-off by an electric motor, controlled by a pressure sensor. In the accumulator, oil pressure is also available when the motor is stopped; it is possible to design the pump according to the average required flow, instead than peak values.

Three different electrovalves are also present in the group:

- One, of proportional type, for clutch actuation
- Two, of proportional type, for the engagement mechanism
- Two, of the on-off type, for the selection mechanism

The choice of valve types for the engagement mechanism is decided by the need to control thrust in order not to damage the synchronizers while obtaining high actuation speeds. The selection valves can be of the on-off type because it is not necessary to control the actuation force with high precision.

The clutch actuator is similar to that already seen for the manual clutch hydraulic mechanism; the shift control system also includes an axial position sensor (engagement stroke) and an angular position sensor (selection stroke). Selected gears can be identified through these sensors.

When the number of selection planes increases, as in a six speed gearbox, the actuation system for the added fourth selection plane could become quite complex; in this case a cam actuation system is used, named *S-cam*, after the shape of its cam.

With reference to Fig. 15.5, the shift actuator is made by a single bistable piston 1 that moves the finger of the internal shift mechanism. The piston can rotate in the cylinder, in order to reach different selection planes.

The two piston stroke ends correspond to the two engagement positions of each sleeve; an intermediate idle position could not be obtained with sufficient precision. Because of this limitation, the internal mechanism is completely reorganized, by putting neutral and reverse on the first selection plane, first and second on the second plane and so on. See the scheme on the lower left of Fig. 15.6.

As can be seen in the same figure, the piston 1 engages through a pin 4 a cam cut on a sleeve 3; this cam is shaped like an S. The sleeve can rotate or be fixed by a magnetic actuator 5. Because there is a spring plunger that can match one of the axial grooves on piston 1 (one groove for each selection plane), piston 1 alternatively engages two neighboring speeds, until actuator 5 is switched on; if

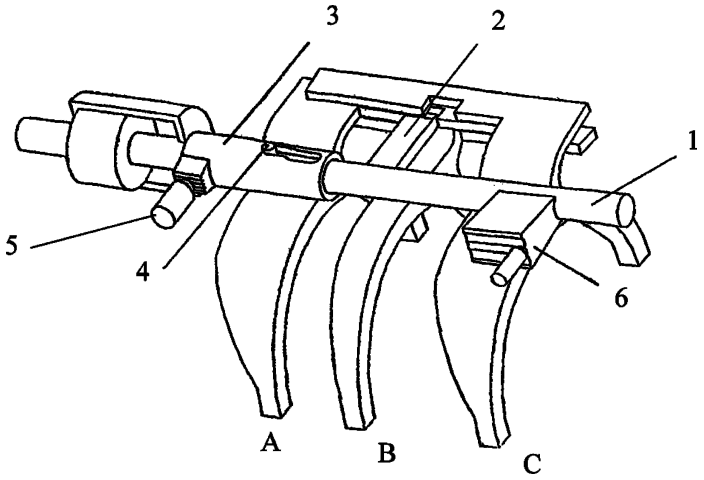


FIGURE 15.5. S-cam actuator (Marelli). See the working details in the next figure.

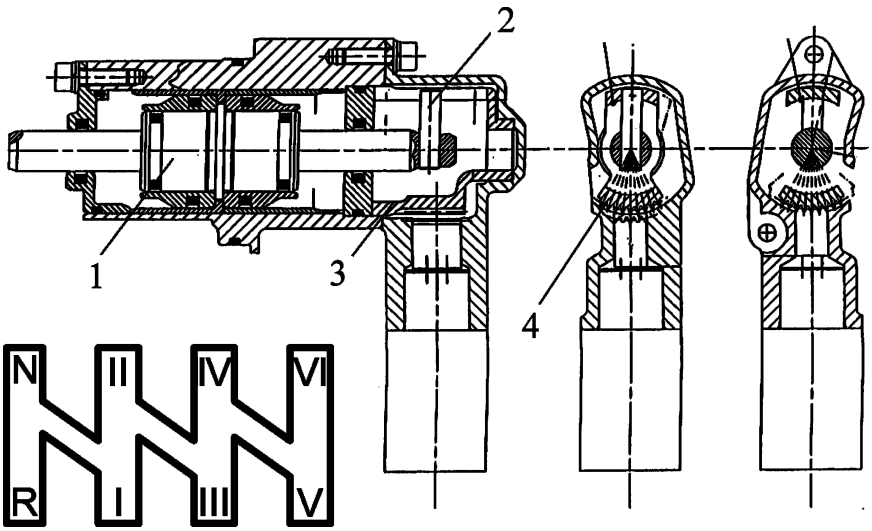


FIGURE 15.6. Working details of an S-cam actuator; on the lower left is the selection scheme of the related gearbox (Marelli).

this occurs, when the piston goes to the left, the finger tip lowers by one selection plane, or rises by one, when the piston moves to the right.

Working on the piston and on the actuator 5 in combination, all speeds can be selected and engaged solely in a sequential way, according to the order R, N, 1, 2, 3, 4, 5, 6 and vice versa.

15.2.2 Multi-disc clutch gearboxes

The configuration of this kind of automatic gearbox is similar to that of a manual gearbox. Synchronizers alone are replaced by multi-disc wet clutches and the speed shift can be of the power-shift type.

Reverse speed is made with an idler, in a way similar to a manual gearbox. With this kind of gearbox there are no limits, as in epicycloidal gearboxes, on choosing the desired transmission ratios.

The simplicity of the basic concept should not suggest that this kind of gearbox can actually be made from a manual gearbox. As a matter of fact, the size of the clutch notably increases the centre distance between the gearbox shafts, while the shafts themselves are complicated by abductions to the different clutches. A gearbox with only five speeds needs three transmission stages.

A construction drawing of a five speed gearbox for a front wheel driven car with transverse engine is shown in Fig. 15.7. In this drawing, the shafts are represented on the same drawing plane; input shaft gears (the first from top) match with those of the output shaft (the second from the top) and with those of a countershaft (the third from the top), through a gear set drafted on the left.

This double gear set is necessary because the rotation of the countershaft must not be inverted when compared with the input shaft; the design complexity is due to limited space.

Let us look, on the upper right of the same figure, at the lock-up torque converter whose hydraulic clutch is fed through a hole on the input shaft; the converter pump moves a gear oil pump for gearbox lubrication and supplies pressure to different actuators.

Gears for third, first and second are on the countershaft, starting from the right. On the input shaft, always from the right, there are gears for fourth and fifth; the radial dimensions of the gears are conditioned by clutch diameters. The reverse speed gears are on the right and are in constant mesh with an idler; a sleeve dog clutch is shared with the fifth speed.

In Fig. 15.8 a singular configuration scheme is reported; it is called *dual-clutch* because two different clutches provide for start-up and power-shift; the scheme is suitable for a front wheel drive with transverse engine, but could be easily adapted for different traction configurations.

Clutches F_1 and F_2 can move two coaxial shafts, where one is fixed to the odd speeds input wheels and the other to the even speeds input wheels. In first speed, for example, clutch F_2 is engaged, while clutch F_1 is disengaged; the second gear shaft is idle and receives no torque. If the sleeve of the second speed is engaged, the internal shaft begins to move but will remain at idle; at any time clutch F_1 can be engaged and clutch F_2 disengaged, according to a powershift procedure.

In the same way first gear can be disengaged, to engage the third, in preparation for the next shift. Speeds must be used sequentially; double up or down shifts are impossible.

The clutch F_2 is also used for starting-up the vehicle. The particular architecture of this gearbox features a reduced number of idle turning clutches and

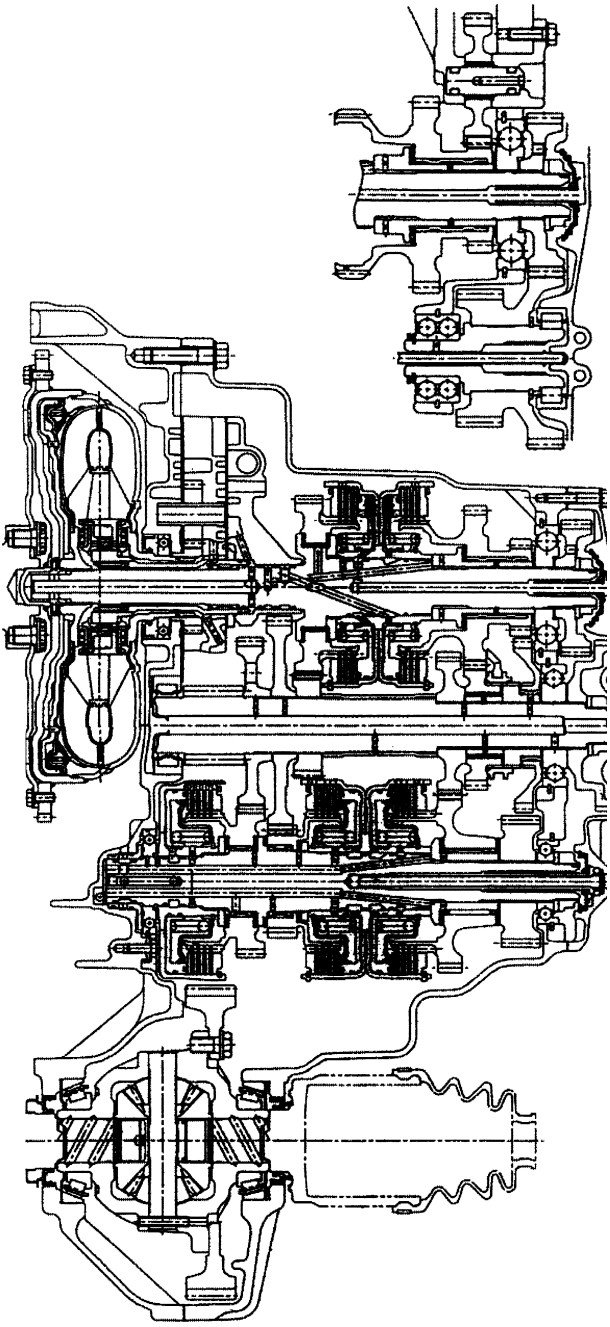


FIGURE 15.7. Countershaft powershift gearbox with multi-disc wet clutches (Honda).

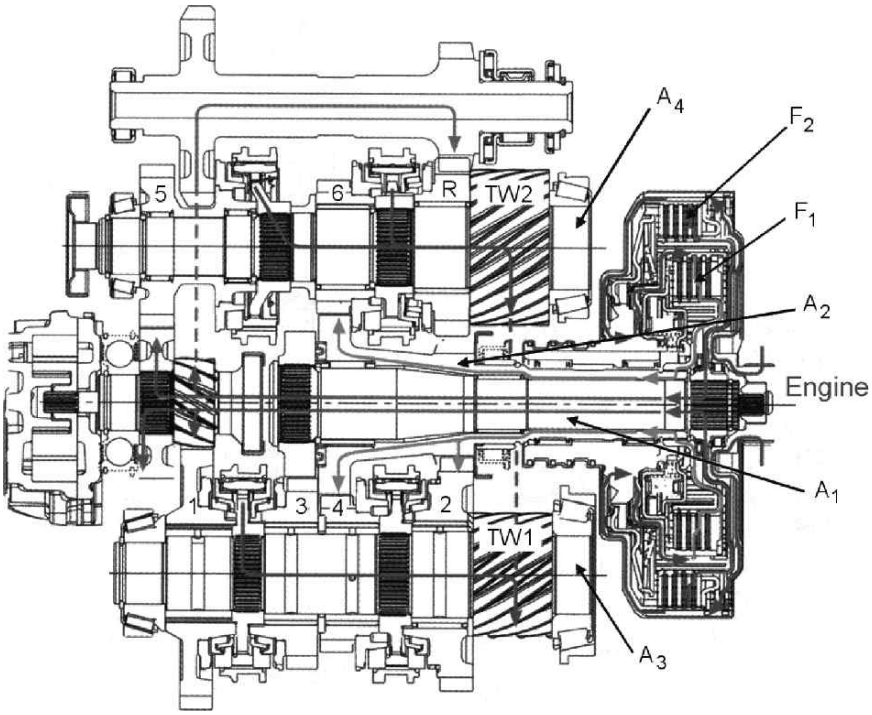


FIGURE 15.8. Working scheme of a powershift dual-clutch six speed gearbox (Audi).

no torque converter. The effect on mechanical efficiency is surely positive; the advantage is offset by an increased design complexity.

The scheme is draft on the same plane, as usual; on the engine shaft are installed the two clutches F_1 and F_2 (start-up device) that move shafts A_1 and A_2; shaft A_1 moves the input wheels of first, third and fifth speeds, while the shaft A_2 moves the input wheels of second, fourth and sixth (a single wheel).

Two countershafts A_3 and A_4 match the final drive wheel and are respectively fixed with the driven wheels of fifth and sixth speeds and with the driven wheels of first, second third and fourth speeds.

The reverse speed shows a third idle countershaft, not visible in the diagram; it can be seen on the front of the following Fig. 15.9. This countershaft meshes with the driving wheel of the first speed on the A_2 shaft and with the R wheel on the A_3 shaft; its synchronizer is adjacent to that of the sixth speed.

15.3 EPICYCLOIDAL CAR GEARBOXES

15.3.1 Epicycloidal trains

We will examine the most important gear trains used on automatic gearboxes; we acknowledge that the configurations in use are more numerous.

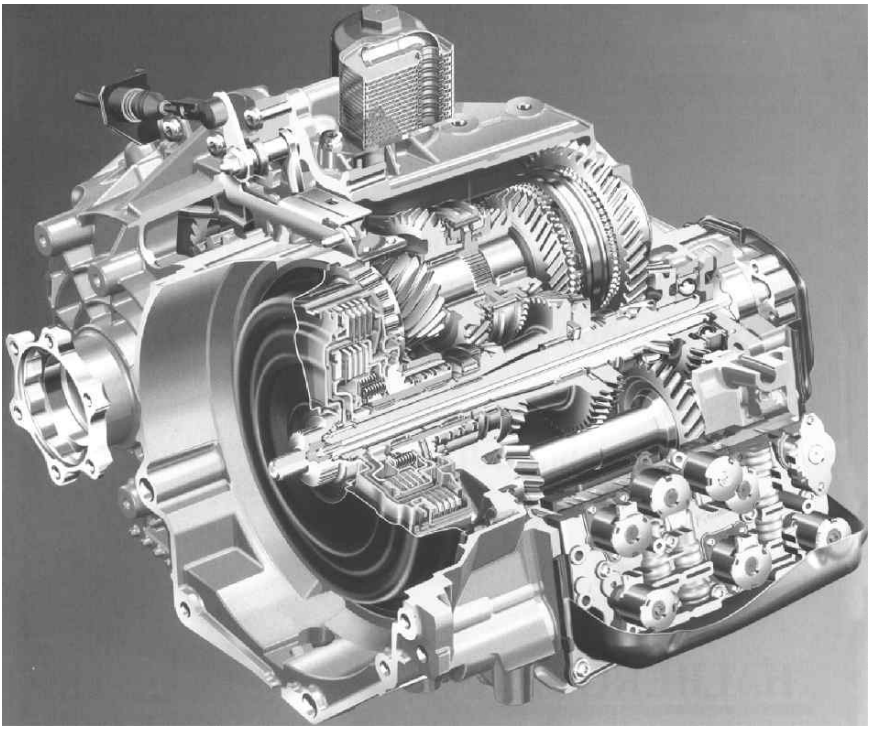


FIGURE 15.9. Cut away view of a dual-clutch powershift six speed gearbox (Audi).

These trains are always made with spur gears; in the past some examples using bevel gears were available.

The simplest configuration for an epicycloidal gear train is shown in Fig. 15.10; the same figure was presented for gear speed reducers, splitters and transfer differentials. The wheel train is made by an internal tooth gear (*annulus*) and an external tooth gear (*sun*) with concentric centre lines; the two wheels mesh with other wheels (*satellites*) whose hubs are supported by shafts mounted on a rotating structure coaxial to the sun, called the *carrier*.

There are normally three satellites affixed to the carrier at constant angles; the radial components of gearing forces are, therefore, self-equilibrated.

These systems show three degrees of freedom; in order to transmit a torque from an input to an output element, the third element must be fixed to the gearbox casing (stator). Three transmission ratios can be obtained with a train of this kind (as many as the elements that can be stator: sun, carrier, annulus), multiplied by two (the number of possibilities for a given element to be input or output shaft). To these six possible transmission ratios, one more should be added as a direct drive, by fixing the input and the output shafts together.

If we designate as:

$$\tau_0 = -\frac{z_3}{z_1}$$

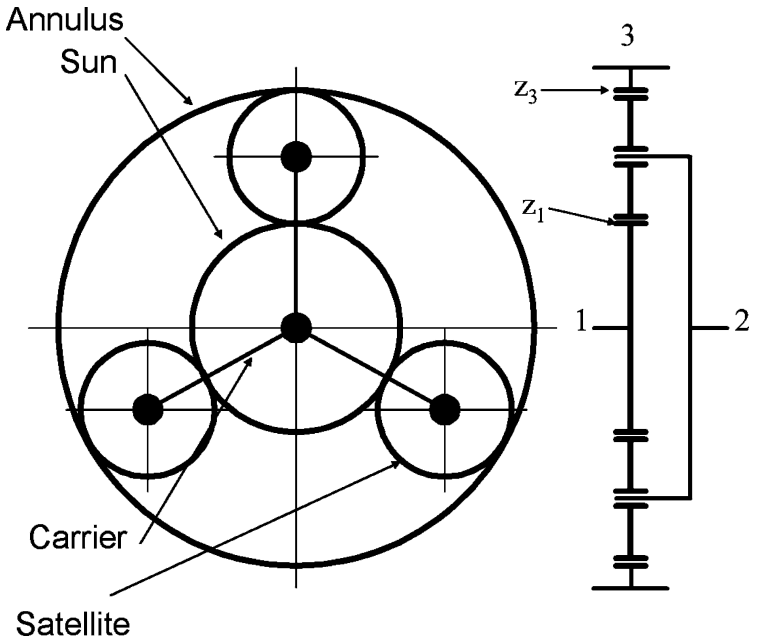


FIGURE 15.10. Scheme of a simple epicycloidal gear train.

TABLE 15.1. Possible values of the transmission ratio of a simple epicycloidal gear, in terms of the constraints and of ordinary transmission ratio τ_0, with reference to Fig. 15.10.

Input	Output	Stator	Transmission ratio τ
1	3	2	$\tau = \Omega_1/\Omega_3 = \tau_o$
3	1	2	$\tau = \Omega_3/\Omega_1 = 1/\tau_o$
1	2	3	$\tau = \Omega_1/\Omega_2 = 1 - \tau_o$
2	1	3	$\tau = \Omega_2/\Omega_1 = 1/(1 - \tau_o)$
2	3	1	$\tau = \Omega_2/\Omega_3 = 1/(1 - 1/\tau_o)$
3	2	1	$\tau = \Omega_3/\Omega_2 = 1 - 1/\tau_o$

the ordinary transmission ratio that can be obtained by fixing the carrier, we can obtain the six different value of the transmission ratio τ reported in Table 15.1.

It should be noticed that τ_0 is negative, because when the carrier is locked, the output shaft is counter rotating as compared with the input shaft; not all ratios can be utilized on the same gearbox, because of the obvious difficulty of having the same element working as input and output in the same mechanism. With this mechanism two ratios (one reduced and one direct drive) can be obtained in practice.

To obtain a higher number of transmission ratios, more epicycloidal gear trains must be combined; an obvious solution is to put more simple gears in

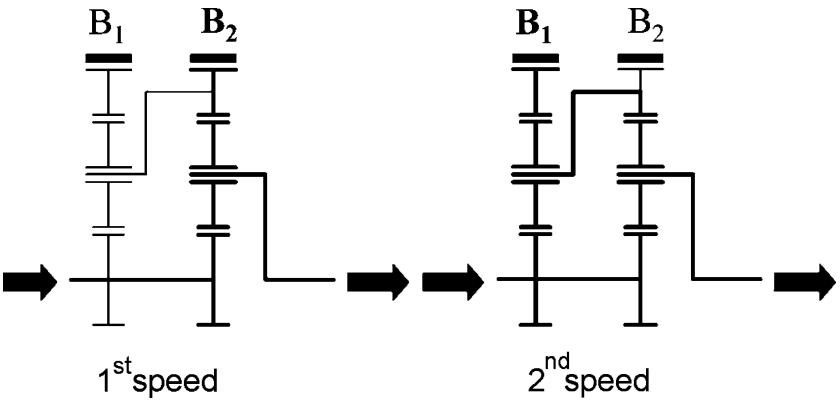


FIGURE 15.11. Scheme of a Wilson compound epicycloidal gear train; two speeds can be obtained by stopping one of the two annulus wheels with B_1 and B_2 band brakes.

series, where each of them makes available one of the necessary ratios alone or in combination with the remaining gears.

A second method is to utilize compound gear trains, where different elements are integrated together. This second way, preferred on somewhat older configurations, allows design simplification but with limited flexibility in determining transmission ratios.

The Wilson epicycloidal gear train, shown in Fig. 15.11, is a first example of a compound train. In this mechanism two simple trains are matched in such a way that the carrier of the first train is fixed to the annulus of the second; by stopping one of the annulus wheels at a time, two different not inverted transmission ratios can be obtained.

By fixing the two carriers together, a direct drive can be obtained.

An alternative scheme of the same mechanism is shown in Fig. 15.12; in this figure B_1 and B_2 are band brakes, while C_1 and C_2 are two multi-disc clutches. Three forward and one reverse speed can be obtained by locking the different elements according to the table shown in the same figure (X designates locked elements).

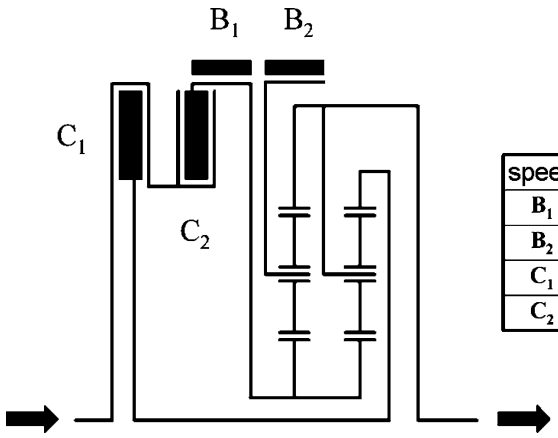
In first speed, when the B_2 brake is locked, we can write for the left mechanism with blocked carrier:

$$\frac{\Omega_{As}}{\Omega_S} = -\frac{z_S}{z_A} \ . \tag{15.1}$$

In the above formula, Ω are the angular speeds, z the gear teeth number, subscripts d and s show the right and left gear train, while A and S stand for annulus and sun gear. It should be noted that teeth numbers show no d or s subscript, because the corresponding wheel of the two simple trains are equal.

For the right train, distinguishing again by the 1 and 2 subscripts the input and output shafts:

$$\frac{\Omega_1 - \Omega_2}{\Omega_S - \Omega_2} = -\frac{z_S}{z_A} \ . \tag{15.2}$$



speed	I	II	III	R
B_1		X		
B_2	X			X
C_1	X	X	X	
C_2			X	X

FIGURE 15.12. Scheme of a second Wilson compound epicycloidal gear train with brakes and clutches; the table shows statuses corresponding to different speeds.

Remember that Ω_{AS} and Ω_2 are equal, because their corresponding elements are fixed together.

From the two equations the first gear transmission ratio can be derived:

$$\tau_1 = 2 + \frac{z_S}{z_A} \ . \tag{15.3}$$

For the second speed the right sun gear is locked and the left simple gear train is idle; the right gear train has the annulus fixed to the input shaft and the carrier fixed to the output shaft. Therefore:

$$\tau_2 = 1 + \frac{z_S}{z_A} \ . \tag{15.4}$$

For the third gear the contemporary lock of clutches C_1 and C_2 puts the mechanism in direct drive, with:

$$\tau_3 = 1. \tag{15.5}$$

For the reverse gear the carrier of the left simple train is locked, while the right train is idle. The sun gear of the left train is fixed to the input shaft and the annulus gear is fixed to the output shaft; therefore:

$$\tau_{RM} = -\frac{z_A}{z_S} \ . \tag{15.6}$$

If, for example, $z_S/z_A = 1/3$, the first speed ratio will be 2.333, the second speed 1.333, the third 1 and the reverse –3; if, instead, $z_S/z_A = 2/5$, the first speed ratio will be 2.4, the second 1.4, again 1 for the third and –2.5 for the reverse speed.

Recalling what we said about torque converters, obtainable ratios are similar to those of a reference manual four speed gearbox (the stall torque converter

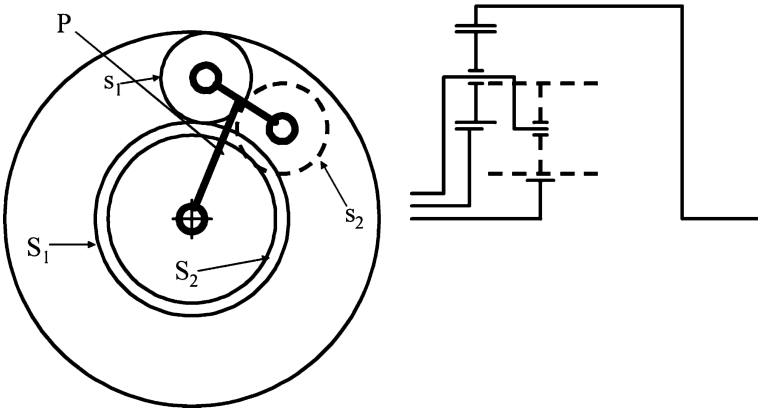


FIGURE 15.13. Scheme of a compound Ravigneaux epicyclical gear train; the satellite s_1 is represented conventionally with dotted lines, because its drawing plane is not on the sketch plane.

torque ratio can be about 2.5); the possibility of obtaining a desired speed step or a suitable reverse speed ratio is unfortunately compromised.

A different compound gear train configuration is that of the so-called Ravigneaux train, shown in Fig. 15.13; it is widely used in automatic gearboxes. In this train a single annulus wheel is provided with two satellites s_1 and s_2 on the same carrier P; the satellites mesh with two different sun wheels S_1 and S_2. Satellites s_2 mesh with the smaller sun S_2; they do not mesh with the annulus but with the satellites s_1 of the larger sun S_1.

Because satellites s_2 have their centre line on a different drawing plane, they are conventionally sketched with dotted lines.

In the following Fig. 15.14, the status of clutches and brakes is shown using the established procedure. Using this scheme, with three multi-disc clutches and two band brakes, four different forward speeds and one reverse speed can be obtained.

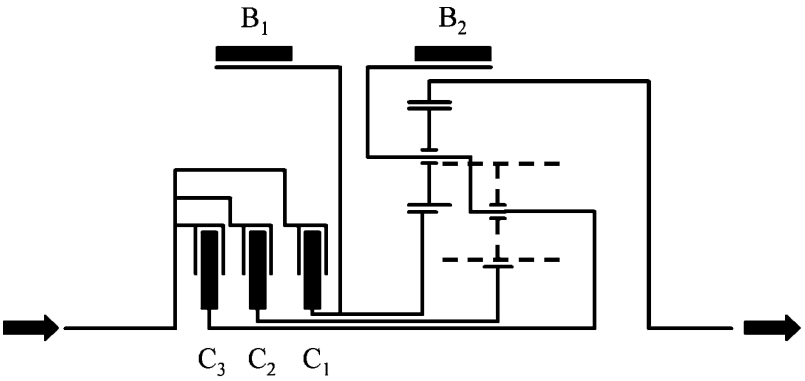
We adopt the symbols of the previous example; we have, in first speed, the right sun gear fixed to the input shaft and the annulus gear fixed to the output shaft, while the carrier is the stator element; therefore:

$$\tau_1 = \frac{z_A}{z_{Sd}}; \tag{15.7}$$

this represents the ordinary transmission ratio of the right gear train; the sign is positive because the double meshing satellites do not change the rotation direction.

In second speed the annulus wheel is fixed to the output shaft, while the input shaft is fixed to the right sun gear, the left being locked. For the left gear train, we can write:

$$\frac{\Omega_{Ss} - \Omega_P}{\Omega_A - \Omega_P} = -\frac{z_A}{z_{Ss}}\ ; \tag{15.8}$$



speed	I	II	III	IV	R
B_1		X		X	
B_2	X				X
C_1			X		X
C_2	X	X			
C_3			X	X	

FIGURE 15.14. Scheme of an automatic gearbox with a Ravigneaux epicycloidal gear train; the table shows the status of clutches and brakes at different speeds.

because the left sun gear is locked and the annulus is fixed to the output shaft:

$$-\frac{\Omega_P}{\Omega_2 - \Omega_P} = -\frac{z_A}{z_{Ss}} \ . \tag{15.9}$$

For the right epicycloidal gear train, we can write:

$$\frac{\Omega_{Sd} - \Omega_P}{\Omega_2 - \Omega_P} = -\frac{z_A}{z_{Sd}} \ . \tag{15.10}$$

If we remember that the right sun gear is fixed to the input shaft, we can also write:

$$\frac{\Omega_1 - \Omega_P}{\Omega_2 - \Omega_P} = \frac{z_A}{z_{Sd}} \ . \tag{15.11}$$

If we compare the equations, we obtain:

$$\tau_2 = \frac{1 + \frac{z_{Ss}}{z_{Sd}}}{1 + \frac{z_{Ss}}{z_A}} \ . \tag{15.12}$$

In third speed the carrier and the right sun gear are fixed, therefore the gearbox is in direct drive, and:

$$\tau_3 = 1; \tag{15.13}$$

in fourth speed, the carrier is fixed to the input shaft and the annulus is fixed to the output shaft, while the left sun gear is locked; therefore:

$$\tau_4 = \frac{1}{1 + \frac{z_{Ss}}{z_A}} \ . \tag{15.14}$$

In reverse speed the carrier is locked, the sun is fixed to the input shaft and the annulus to the output shaft; the transmission ratio is that of the ordinary left gear train and therefore:

$$\tau_{RM} = -\frac{z_A}{z_{Ss}} \ . \tag{15.15}$$

Remarks made for the previous case about the possibility of obtaining a given range and ratio step still apply.

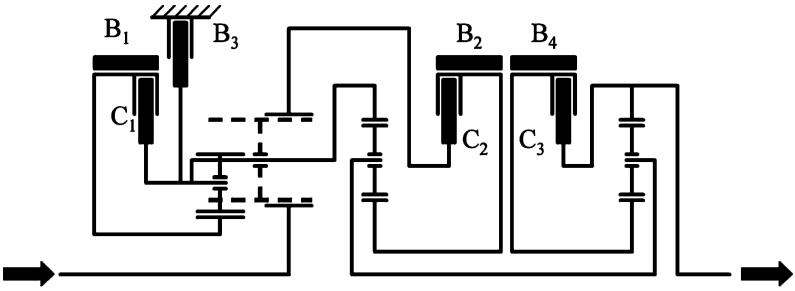
If the number of available speeds increases, the applied gear trains becomes more complex; for example a scheme suitable for a five speed gearbox may be conceived by adding a simple epicycloidal gear train to the Ravigneaux gear train previously described.

The scheme could be compared to the application of range speed reducers that we described for industrial vehicles.

The total number of speeds should be eight, by multiplying the four speeds of the compound gear train by the two of the added speed reducer (reduced and direct drive); in reality many obtainable speed ratios would be too close to each other and therefore useless.

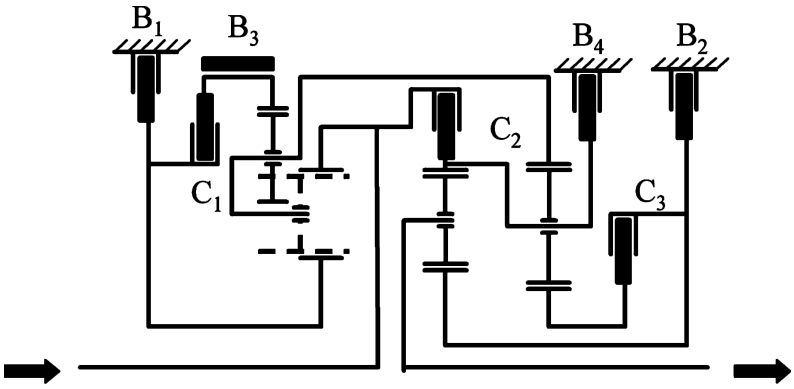
The simplest scheme, considered until recently, has been abandoned in favour of compound gear trains that look unnecessarily complicated, but obtain transmission ratio values closer to theoretical specification.

Figure 15.15 shows the example of a Ravigneaux compound gear train joined to two simple epicycloidal gear trains; the number of theoretically available speeds is well over the five practically implemented. Because many possible



Speed	I	II	III	IV	V	RM
C_1			X	X	X	
C_2	X			X	X	X
C_3	X	X	X	X		X
B_1		X				
B_2	X	X	X			
B_3						X
B_4					X	

FIGURE 15.15. Scheme of an automatic powershift gearbox featuring three epicycloidal gear trains; five different speeds are obtained (Mercedes).



Speed	I	II	III	IV	V	VI	VII	RM_1	RM_2
C_1			X	X	X				
C_2				X	X	X	X		
C_3	X	X	X		X	X	X	X	X
B_1		X				X			X
B_2	X	X	X	X					
B_3	X						X	X	
B_4								X	X

FIGURE 15.16. A scheme derived from the previous; it allows seven forward speeds and two reverse speeds (Mercedes).

combinations were discarded, the implemented transmission ratios come close to assuming the desired values and therefore allow a better optimization of performance, fuel consumption and emission.

In the same figure is represented, using the same procedure, the status of brakes and clutches for different speeds; this scheme has been designed for and applied to conventional driven cars with longitudinal engines.

Quite recently, this scheme has been modified in a very simple manner, to obtain what is represented in Fig. 15.16; essentially an additional annulus wheel has been added to the Ravigneaux gear train and the lay-out of brakes and clutches has been modified, without increasing their number. The result is seven forward speeds and two reverse speeds, as the status table shows in the picture.

The transmission ratio calculation will be omitted for the two previous schemes for the sake of simplicity.

This overview on gearbox schemes does not pretend to be a complete report on what is or can be produced. Possible combinations of simple and compound epicycloidal gear trains are numerous.

The epicycloidal gear solution is, for the time being, the most widely used for powershift automatic gearboxes.

15.3.2 Production examples

Conventionally driven cars benefit from the alignment of input and output shafts.

The example in Fig. 15.17 corresponds to the scheme in Fig. 15.15 and adopts, as we have seen, a Ravigneaux gear train 3 and two epicycloidal speed reducers 2 and 4, downstream of the first; it features five forward speeds and a reverse speed.

We note that on torque converter automatic gearboxes it is impossible to keep the vehicle braked while parking, by engaging only the lowest speed, as is possible on manual gearboxes.

On manual gearboxes engine compression and friction torque are used to withstand the tendency of the vehicle to move. The torque converter, inserted on the transmission line, acts as a free wheel joint at low speed; the lock-up clutch, if any, could not keep the car braked, because actuation pressure is not indefinitely available when the engine is stopped.

To keep the vehicle stopped, a manually actuated park pawl is added, matching a sprocket wheel on the output shaft; in the figure the pawl is the element 15 and the corresponding sprocket wheel is the element 14.

The torque converter can also be seen (without lock-up clutch), as can the oil pump 6 and the different multi-disc wet clutches used for gear engagement.

A recent evolution of this scheme is represented in Fig. 15.18; the operating principles are explained in the previous Fig. 15.16. The epicycloidal gear trains are similar to the previous case. A lock-up clutch can be noticed on the torque converter.

Automatic gearboxes for front wheel driven cars, with longitudinal or transversal engine, are complicated by the necessity of having an output shaft at a convenient distance from the input shaft, not coaxial as on epicycloidal gearboxes; in addition, space limitations caused by the front steered wheel are particularly critical on transversal engines.

Nevertheless, automatic gearboxes have gear trains that are substantially similar to those of conventional drives. Additional single stage trains are added to locate the output shaft at the desired final drive pinion position.

One example of this architecture is reported in Fig. 15.19, for a front wheel driven car with longitudinal engine.

The scheme includes both a simple and Ravigneaux gear trains that in conjunction with three clutches and two brakes make six speeds available. The Ravigneaux gear train output shaft moves the differential final drive through an auxiliary shaft. The presence of the all wheel drive does not influence the lay-out.

The possibilities available to front wheel driven cars with transversal engine are two.

A first solution provides for torque converter and oil pump in line with the engine; the torque converter output shaft is fixed to a sprocket wheel that, through a silent chain, moves the input shaft of the real gearbox on a parallel

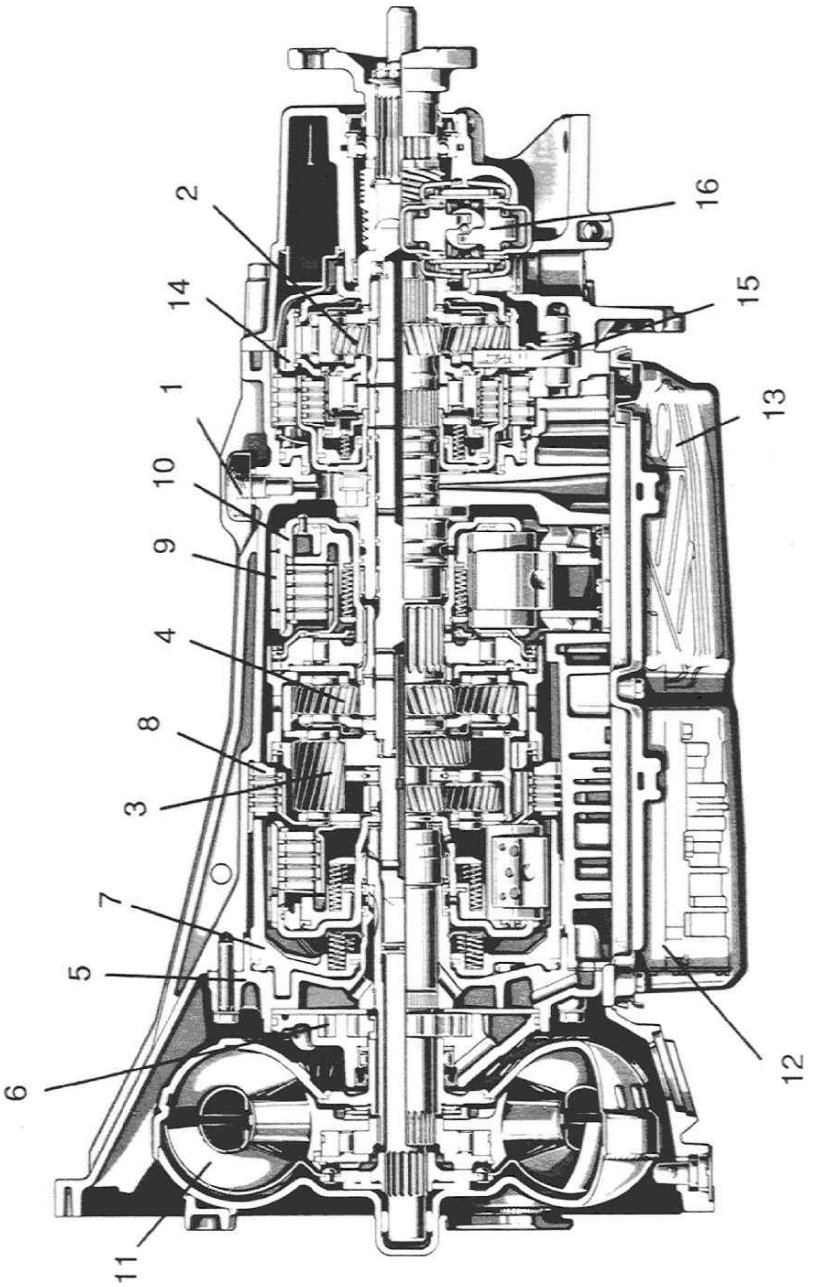


FIGURE 15.17. Scheme for a five speed automatic powershift gearbox for conventional drive (Mercedes).

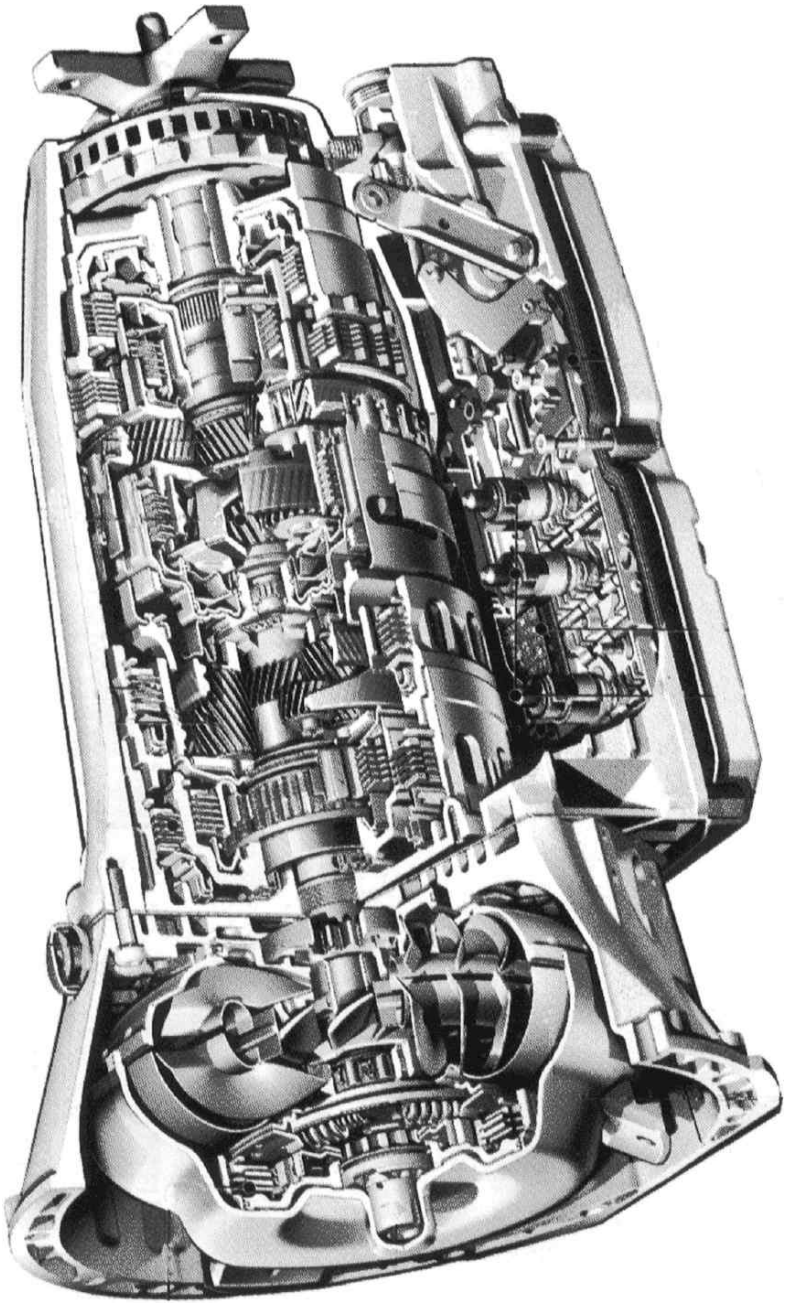


FIGURE 15.18. Scheme for a seven speed automatic powershift gearbox for a conventional drive (Mercedes).

FIGURE 15.19. Six speed automatic gearbox with epicycloidal gear trains; the lateral shaft connects the gearbox output shaft with the differential pinion, behind the engine (Audi).

centre line. Every geometrical problem regarding the transversal direction of the car is eliminated; on the other hand consequences for the front overhang of the car become relevant.

A second solution divides the epicycloidal gearbox into two sections.

The first section could include a Ravigneaux gear train coaxial to the engine; its output shaft moves through a single stage train a simple epicycloidal gear train coaxial to the differential pinion. It is therefore possible to have five speeds within the space available on a front wheel drive with transversal engine.

15.4 CAR CVTs

15.4.1 Motivations

Continuously variable transmissions (CVTs) are mostly found on front wheel drives. They are particularly appreciated for making available an unlimited number of transmission ratios, with benefits for comfort, performance and fuel economy.

A simple explanation of this sentence can be supported by Fig. 15.20. In this figure the specific fuel consumption map is shown using engine torque and speed as parameters; for each output power value, one and only one minimum fuel consumption working exists.

These points are represented by the thick line below the wide open throttle curve; it can be drawn by connecting all tangency points of equal fuel consump-

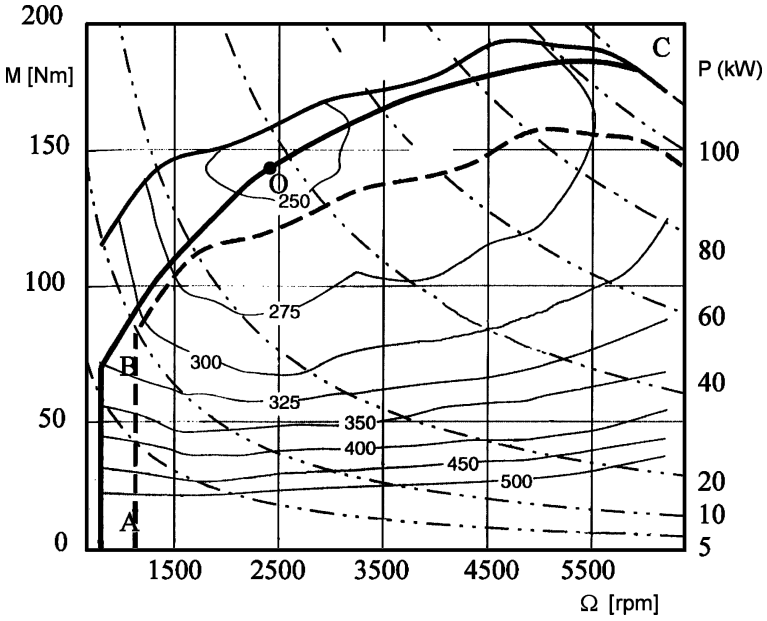


FIGURE 15.20. Engine specific fuel consumption map; the minimum fuel consumption line ABOC and the area of practical use for a stepped gearbox are shown. This area is included between the wide open throttle curve and the dotted curve showing the torque, when existing, obtained at the higher speed.

tion curves (quoted in the figure) with the hyperbola (drafted with a line and dots) representing the constant power curves.

This curve is made by a vertical straight line AB on the left of the map that represents the minimal stable working speed of the engine; this speed depends upon engine characteristics, on the elastic and damping properties of the transmission line (made of clutch and gearbox), on car suspensions and on tires.

Other notable points of this curve are point C, coincident, by definition, with the full power point, and point O that represents the minimum absolute fuel consumption value.

Let us assume changing coordinates in such a way that they represent wheel torque and vehicle speed (instead of engine torque and speed) at the actual speed $i-1$ of a stepped gearbox; the dotted curve, similar to the maximum torque curve, represents the wheel torque at the next i speed.

A very efficient usage of a stepped gearbox allows specific fuel consumption to fluctuate between the maximum torque curve at the $i-1$ speed and the maximum torque curve at the i speed, at a given power level; actual fuel consumption can be higher than the minimum.

At the last speed, fuel consumption can fluctuate between the maximum torque and the road load torque at the last speed; in this case the fuel consumption increase could be substantial.

It is possible to have the engine working at the specific minimum fuel consumption only if an unlimited number of speeds is available.

This fact explains why in recent years so many efforts have been made to increase the number of speeds either for manual or automatic gearboxes.

It should be said that only on the latter, when a sufficiently sophisticated control system is available, is the instantaneously optimum speed guaranteed to be engaged.

Similar statements can be made about acceleration and pollutant emissions.

Possible systems suitable for a CVT include in principle:

- Electric transmissions
- Hydrodynamic transmissions (among them the already known torque converter)
- Hydrostatic transmissions
- Variable geometry mechanical transmissions

The first three categories have been frequently considered in the past, but have been abandoned on conventional vehicles because of poor transmission efficiency; this penalty jeopardizes the advantage theoretically available to a CVT.

These kinds of transmissions can be reconsidered for hybrid vehicles, where the intermediate form of energy generated by the transmission (electric or hydraulic pressure energy) can be used to recover vehicle kinetic energy or used as a temporary alternative to fossil fuel.

Variable geometry mechanical transmissions can show efficiency similar to that of gear transmissions. At this time belt or chain drives are available and roller traction drives are at the developmental stage.

In special purpose vehicles, such as tractors, earth moving machines, airport boarding buses, etc. some hydrostatic CVTs are used.

15.4.2 Production examples

Steel belt transmissions represent the most widely used technical solution for CVT. Steel belts that are known today are the Van Doorne belt and the LuK chain, which is also used as a belt. Their working principles present many common points but some differences in terms component design; this issue will be explained later.

The cross section of a CVT with a Van Doorn steel belt is shown in Fig. 15.21 with the usual drawing system. The gearbox in the figure does not include a

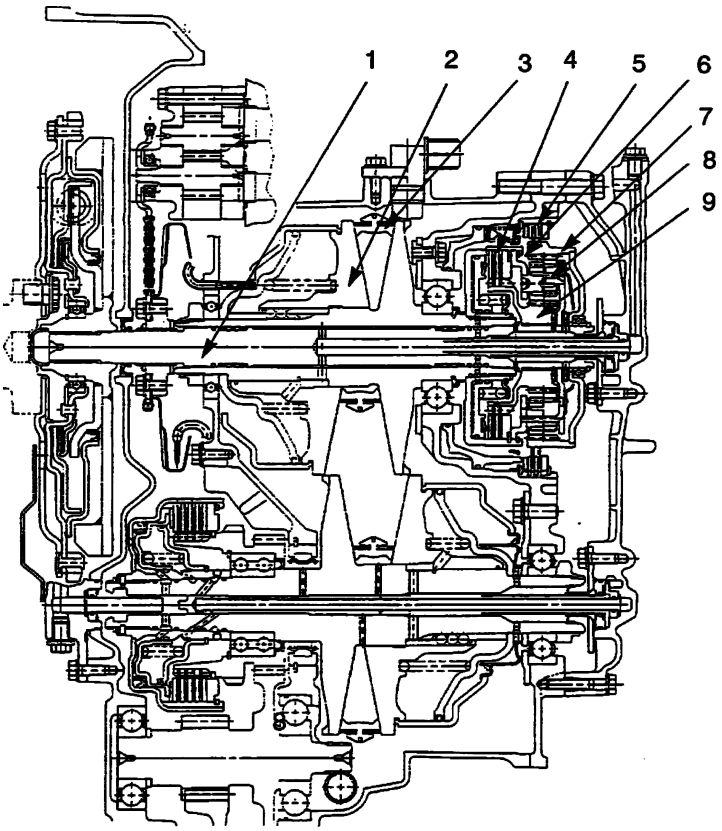


FIGURE 15.21. Van Doorne steel belt CVT; the gearbox is characterized by the start-up multi-disc clutch on the output shaft (Honda).

torque converter, because start-up functions are performed by an electronically controlled multi-disc clutch on the output shaft.

The reason for this choice is two-fold:

- While towing the car, any damage on the belt is avoided; the belt in fact is loose, without oil pressure on pulleys.
- During emergency braking manoeuvres, the transmission ratio can be changed independent of vehicle speed, making the next acceleration quicker.

The engine is fixed to the carrier 8 of an epicycloidal gear train with twin satellites that are used for forward and reverse speed; the gearbox input shaft 1 is instead fixed to the sun gear 9 of the same train. By using the clutch 4 it is possible to set the train in direct drive; by engaging instead the brake 5, the annulus is locked and the sun gear rotation is opposite to that of the carrier, obtaining reverse speed.

With this design it would be possible, if necessary, to change transmission ratio in both forward and reverse drives.

The transmission ratio variation is obtained through two variable gauge V pulleys that match the belt (see detail in Fig. 15.22).

Each pulley is made by two steel cones that can slide axially but are fixed together when turning. Looking at the drawing, the left input semi-pulley and the right output semi-pulley are also fixed in the axial direction; the other two semi-pulleys are instead free to slide.

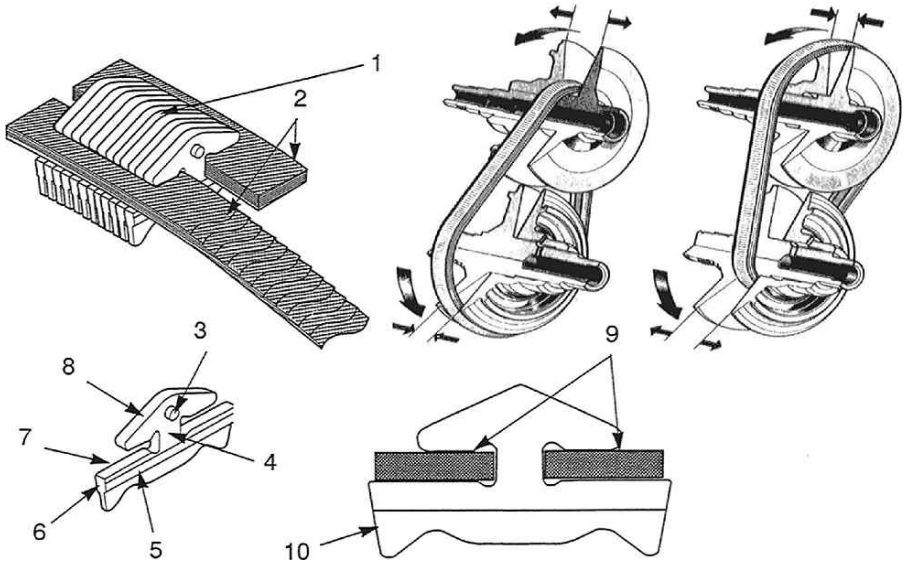


FIGURE 15.22. Detail of the pulley motion in a Van Doorne steel belt CVT (upper part); detail of the belt (lower side). The torque is transmitted by belt compression.

Because the developed belt length cannot change, by narrowing the two driving semi-pulleys the driving pulley primitive diameter will be increased and the two driven semi-pulleys must be accordingly widened, to reduce their primitive diameter; the belt will be shifted to the right and the transmission ratio will be decreased. With the opposite motion the transmission ratio will be increased.

A given transmission ratio range will be obtained between a minimum overdrive ratio and a maximum reduced ratio.

The two limit ratios are reciprocal numbers; this situation, where the 1:1 value is in the middle of the ratio range must be adapted to vehicle use by an additional final drive that can bring the higher ratio close to one.

For this reason there is an idler gear final drive between the output pulley and the differential pinion; it will be necessary, in any case, to position different centre lines appropriate to the engine compartment lay-out. This final drive is also used to adapt the same gearbox to different cars.

The active surfaces of the semi-pulleys are moved by hydraulic pistons; this motion can change their primitive diameter. Oil pressure determines the contact force of the semi-pulley with the belt.

Controlling this parameter is vitally important, because contact force determines torque transmission through friction between pulley and belt; if this force is insufficient, parts would slip, causing loss of efficiency and surface damage. If this force is excessive, mechanical losses would also be higher than necessary, because the contact area between belt and pulley exceeds the kinematically correct primitive circumference.

Van Doorne belts use many thrust elements 1 such as those on the lower side of Fig. 15.22.

Each of these elements contacts the next and is set along the belt centre line; the elements are kept aligned with a series of flexible steel ring bands 2, one inserted inside the other with no clearance. Thrust elements have a trapezoidal shape with a median appendix; the two inclined sides 10 match the sides of the pulleys and exchange the friction force necessary for torque transmission.

The median appendix 4 of each thrust element matches the two sets of concentric rings 9; the contact surfaces of the rings are inclined in such a way that the rings do not come in contact with the sides of the pulley.

The rings must be very thin to contain the curvature stress within reasonable values; the necessary tension is supported by an adequate number of rings.

The extended length of the ring must allow an assembly without clearance to distribute tension between the rings uniformly.

Thrust elements can only exchange compression forces with neighboring elements; therefore the Van Doorne belt will never work with positive tension and will transmit the torque through the compression of the driven part of the belt circuit.

The sole purpose of the rings is to keep the elements at the correct radial position in the pulleys, and to keep them aligned when outside of the pulleys.

Friction forces between elements and pulleys would waste no energy if there were no relative motions. These forces are limited because there is no macroscopic

slip between parts; nevertheless, there is local microscopic relative motion because of two facts:

- The thrust element contact area must have a certain radial extension to limit contact pressure; at each contact area there will be a small slip on points outside the primitive radius.
- Thrust elements entering and leaving the pulleys must slide because they change their path from straight to circular.

The power wasted by these small relative motions will be in any case determined by pressure acting between the belt sides and pulleys; this pressure must be limited to the minimum necessary to avoid macroscopic slip. Contact pressure must therefore be carefully adjusted as a function of transmitted torque.

There are also belt CVTs with a torque converter between engine and speed variation unit.

Other CVTs apply different steel belts that would be better termed chains.

Figure 15.23 shows a chain CVT lay-out, not dissimilar from the first as far as the pulleys are concerned; start-up clutch has been instead positioned before the speed variation unit. This gearbox shows the usual final drive before the differential drive, but the differential shows a bevel gear drive, to adapt the gearbox to a longitudinal engine.

The LuK chain is represented in the upper part of the same figure.

The links of this chain are made by a stack of plates mounted on pins, rolling on their contacts without slip; the pin ends are inclined to match the pulley sides. The friction between pin tips and pulley transmits forces from belt to pulleys and vice versa; in this case the belt is in tension on the driving side of its circuit.

Plate length can vary in different links, to avoid whistles and vibrations.

A positive feature of the chain is the possibility of adopting a smaller minimum radius on the pulley, reducing dimensions at the same level of performance; mechanical efficiency should also be slightly improved.

Chain or belt design have presented many problems; steel has been chosen as most suitable material because of the level of stress and for reasons of oil compatibility.

Many solutions using reinforced elastomeric materials have been also proposed for belts; these must be used outside the lubricated areas of the gearbox and have found their application in situations of limited torque, particularly on scooters.

A different working principle usable for CVT, not yet found in mass production, is that of rolling bodies able to exchange significant tangent forces with rather limited contact areas. This is possible thanks to special synthetic lubrication oils, which have the property of increasing their viscosity with contact pressure (*traction fluids*).

The advantage of this principle consist in its high mechanical efficiency, which can be attained because of limited slips between the contacting parts,

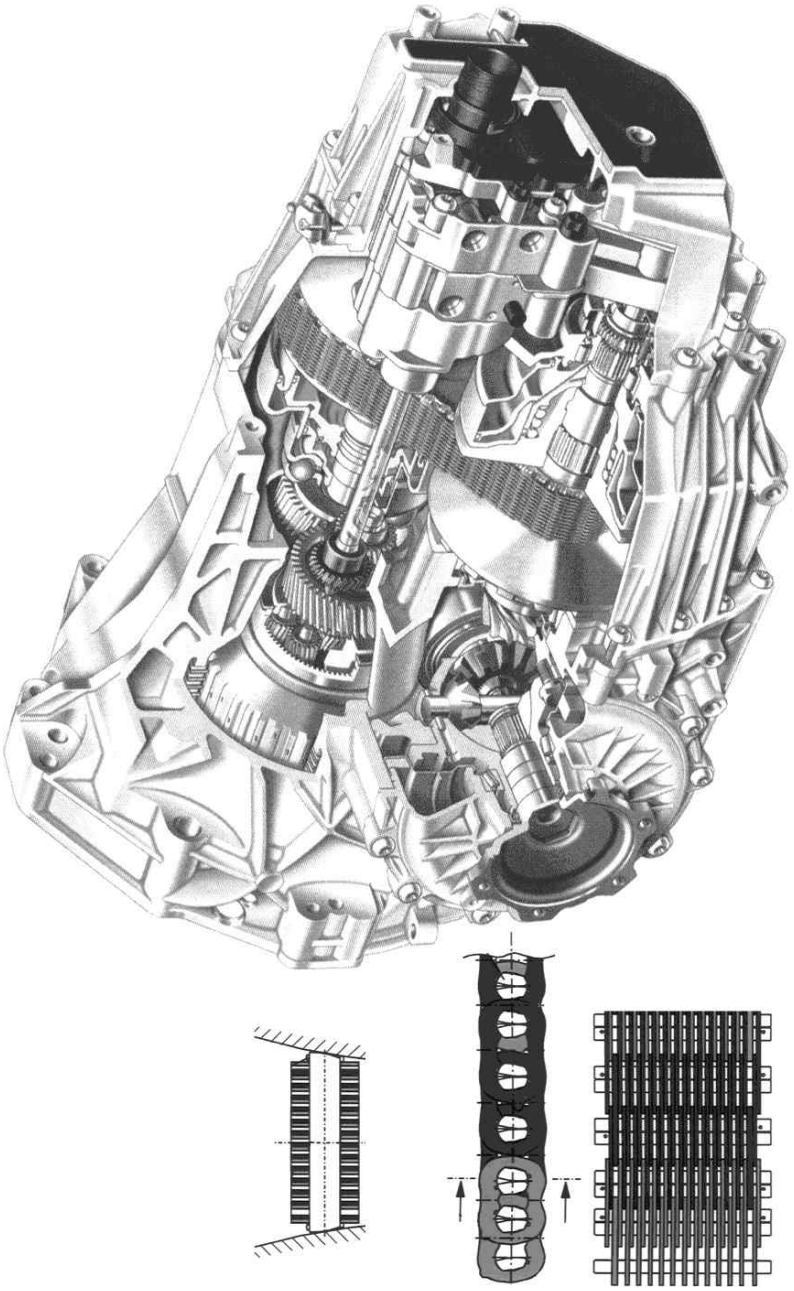


FIGURE 15.23. On the top: chain CVT suitable to front wheel drive with longitudinal engine (Audi). On the bottom: view and cross section of a LuK chain.

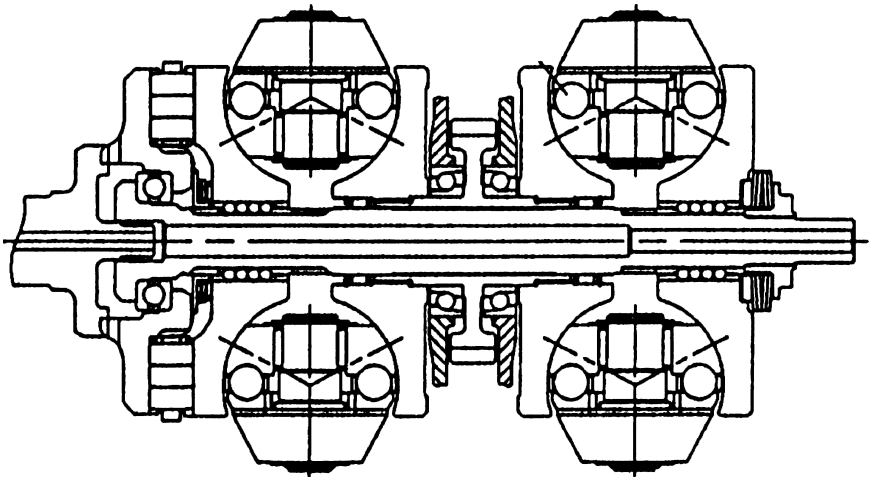


FIGURE 15.24. Rolling bodies toroidal CVT; by steering satellites appropriately it is possible to change transmission ratio between input (on the left) and output pinion (in the middle) (Torotrak).

working very close to pure rolling conditions; the disadvantage is due to the high weight of the mechanism, necessary for obtaining a sufficient contact pressure to produce adequate traction at lubricated conditions.

A typical configuration according to this principle is that of the so called toroidal gearbox, the essential parts of which are shown in Fig. 15.24. The gearbox shows two identical traction bodies, each of them made by two symmetric toroidal surfaces.

These toroidal surfaces are made by two revolution bodies obtained by turning a circumferential arc around the rotation axis; on rotating facing surfaces, a number of rollers are set whose main curvature is slightly higher than that of the toroidal surface.

By changing the angular position of the rotation axis of these rollers, it is possible to change the transmission ratio between toroidal bodies; when the rotation axis of the rollers is perpendicular to the toroidal surface rotation axis, the transmission ratio will be 1:1. By turning the rotation axis in the two directions on the same plane it will be possible to decrease or to increase the transmission ratio.

In order to limit reaction loads on the casing, two symmetric mechanisms are employed whose thrusts can be self-equilibrated in a simple way though a central shaft. The output shaft meshes with the central pinion fixed to the two toroidal driven surfaces.

Three rollers are employed to limit contact forces; a particular mechanism must be used to obtain the same centre line inclination on all of these rollers.

15.5 GEARBOXES FOR INDUSTRIAL VEHICLES

Automatic gearboxes for industrial vehicles show no substantial differences from those of automobile gearboxes; the countershaft architecture is widely used on semi-automatic gearboxes or in automatic gearboxes with many speeds, where an epicycloidal train configuration could be difficult to adopt. In these applications both clutches and torque converters are used.

Epicycloidal gear train gearboxes, no different from those considered on cars, are particularly applied on buses, where comfort is a major priority and the number of speeds is limited.

A particular problem for industrial vehicles is torque converter behavior. The mechanical efficiency of this device is particularly poor when pump and turbine exchange functions; this happens when reaction torque inverts because the engine is expected to brake the vehicle with its mechanical friction or with dedicated devices (*engine brake*).

This fact results from blade angles that cannot be optimized for two opposite flow directions.

Engine brake effect at these conditions is particularly decreased, as compared with a manual gearbox. This fact can be accepted and is sometimes beneficial for fuel economy in cars, but is unacceptable on a heavy vehicle.

Some specific devices (*retarder*) are used to solve this problem; these will be explained later.

15.5.1 Semi-automatic gearboxes

Together with the already discussed pre-selection semi-automatic gearboxes, automatic and automatized gearboxes are on the market with clutch and torque converter in series.

The torque converter is first set with the purpose of increasing start-up torque and smoothing transmission output torque; the torque converter features a lock-up clutch to increase transmission efficiency in cruise drive.

Torque converters include an additional free wheel between pump and turbine; this free wheel is designed to be able to transmit only negative torque (i.e. it is like an open joint when the engine is driving, and like a locked joint when the engine is braking); in this way the desired engine braking effect is achieved.

After the torque converter, an additional component is fit, and not only because of gearbox automation; this is the retarder that helps the vehicle brakes on long downhill drives.

Retarders can be assimilated to a high diameter hydrodynamic clutch, with a limited radial blade dimension; this shape is dictated by the need to limit longitudinal gearbox length and obtain a high stall torque. See part number 7 on Fig. 15.25.

The pump is fixed to the torque converter output shaft; the turbine blades are cut directly on the gearbox casing.

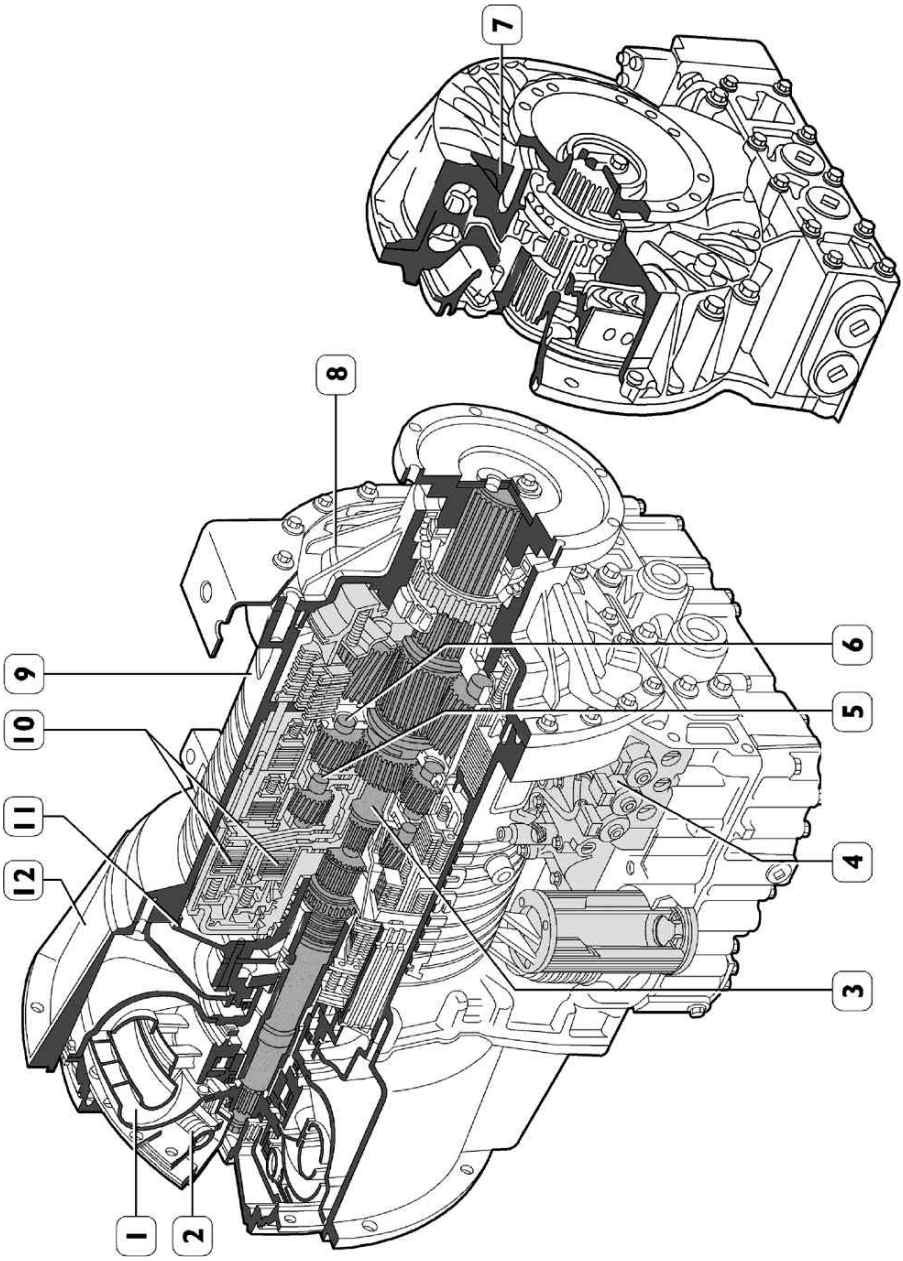


FIGURE 15.25. Epicycloidal gear train automatic gearbox for industrial vehicle; on the right the retarder unit 7 is shown. It can be optionally mounted in place of the back cover 8 (Iveco).

During vehicle slow-down, the hydrodynamic clutch absorbs a braking torque equal to stall torque and dissipates the corresponding power as heat; to adjust or delete braking torque, the quantity of clutch oil is adjusted or eliminated with dedicated valves.

The remaining gearbox elements are those already seen.

15.5.2 Automatic gearboxes

Totally automatic gearboxes for industrial vehicles are also available. These are almost identical to those of conventional drive cars with torque converter and epicycloidal gear trains.

The number of speeds is limited to 5 or 6; this kind of gearbox is therefore suitable for buses or limited speed trucks, subject to frequent stops.

Figure 15.25 shows an example of this gearbox, with six forward and one reverse speeds; three simple epicycloidal gear trains can be used separately or in combination.

The back gearbox cover 8 can be changed with the retarder unit 7 we described in the previous paragraph. In this case the radial dimensions are reduced by using a two sided bladed wheel; the central rotor is fixed to the output shaft, while the stator is cut in the casing.

Regulation of braking torque is managed by the quantity of oil in the clutch.

We introduce finally a particular automatic gearbox suitable for urban buses; it was developed to obtain different retarded speeds useful in hilly towns.

The gearbox is presented in Fig. 15.26 and is characterized by having the torque converter installed in the central section of the gearbox, instead on the joint flange with the engine, as usual.

The torque converter (6 is the pump, 7 is the turbine, 8 the stator) is used as a start-up device as well as a retarder; input shaft 1 is fixed to the engine through the joint 2 that acts as a torque damper. Two different clutches 3 and 4 are provided for a first stage epicycloidal gear train 13 and for the direct drive. The two output epicycloidal gear trains 17 and 18 are used to mix energy flows through the input shaft and the turbine and for the reverse speed. The heat exchanger 19 can dissipate the heat generated by the retarder.

The scheme in Fig. 15.27 and the related status table of brakes and clutches allow us to understand how the mechanism works.

In first speed, power flows through two parallel channels. The first channel is made by a rigid connection with the epicycloidal gear train carrier 13; the second channel flows through the torque converter, whose pump is fixed to the sun gear of the same train. The turbine is fixed to the sun gear on train 17. The annulus gear is fixed to the input shaft.

In second speed, the revolution speed is reduced by the epicycloidal gear train 13 whose sun gear is locked; the torque converter is not loaded.

In third speed the gearbox is a direct drive.

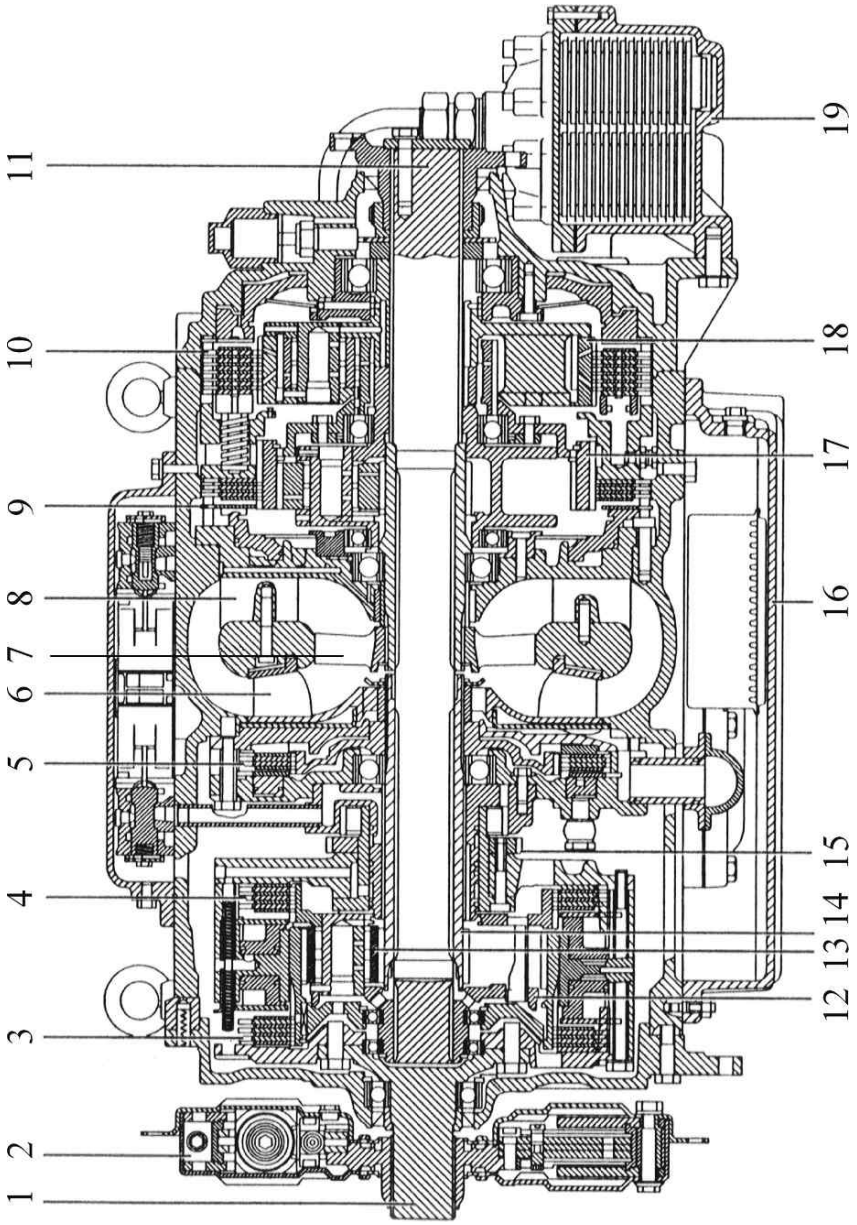
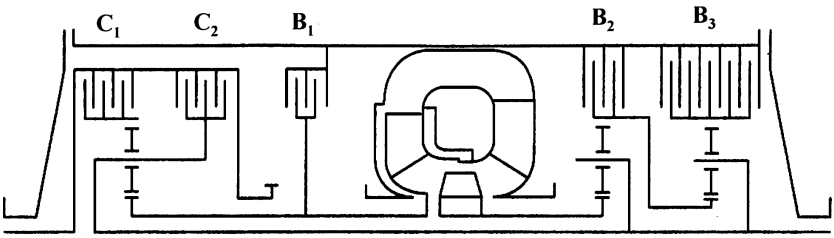


FIGURE 15.26. Automatic gearbox for urban bus. The architecture is characterized by a torque converter in the central section of the gearbox; it is used for vehicle start-up as well as a retarder (Iveco).



Speed	I	II	III	I R	II R	III R	RM
C_1	X	X			X		X
C_2			X			X	
B_1		X	X	X		X	
B_2	X						
B_3				X	X	X	X

FIGURE 15.27. Working scheme of the previous automatic gearbox; three normal forward speeds, three retarded forward speeds (I R, II R, III R) and one reverse speed are available.

In retarded speeds the third brake involves the epicycloidal gear train 18 that rotates the turbine in the opposite direction. The turbine acts now as pump; the pump is locked by the B_1 brake and acts as a stator.

The torque converter is now an effective retarder.

15.6 CONTROL STRATEGIES

The following paragraphs will explain the main rules for speed choice and speed shift actuation that automatic gearbox control systems must observe to interpret the driver's intentions in the best way, taking into account the power requested by road profile and speed.

We assume that the gearbox adopts an electrohydraulic control system where engagements and disengagements are made by hydraulic actuators fed through electronically controlled electrovalves. This technology is common to new automatic gearboxes, even if some applications exist on the market where also control functions are also performed by hydraulic logic circuits.

A typical electronic control system for an automatic powershift gearbox with epicycloidal gear trains receives the following magnitudes as inputs:

- Speed selector position; typically positions P, R, N, D exist indicating situations for parking, reverse speed, neutral position and automatic forward drive; selectors can also show:

  - Fixed speed positions (i.e. 1, 2, 3, 4) to indicate driver's will to lock the gearbox in one of the speeds or, alternatively, to limit automatic shifts up to one of these speeds
  - Selectors to choose different automation programs such as W for winter (when ice and snow are on the road the first speed is inhibited and shifts are slower), E for economy to show the driver's desire to drive at low fuel consumption, S for sport driving, etc

- +/– Positions to use the gearbox sequentially, increasing or decreasing speeds of one unit at a time; in this case additional commands may exist on the steering wheel.
- Engine speed.
- Throttle position in petrol engines or pumped quantity on diesel engines. More recent engines may have a drive-by-wire regulation system: Accelerator pedal position or some calculated entity may indicate the required power.
- Cruise control position.
- Torque converter turbine speed.
- Gearbox output shaft speed.
- Accelerator kick-down position.
- Oil brake pressure or stop light switch, indicating a braking situation.
- Gearbox oil temperature; this sensor is particularly important because oil actuator response is influenced by viscosity and therefore temperature.
- Engine coolant temperature, showing the readiness of the engine to deliver maximum performance.
- Driver's door status (open or closed) to inhibit gear selection from the outside. Other safety signals may also be used.

The control system normally manages the following output actuators:

- On/off electrovalves, as many as hydraulic lines for brake and clutch actuation; valve number may be reduced by combining more functions on the same valve.
- Proportional valves, as many as the functions to be simultaneously actuated with regulated pressure; at least one for the next speed engagement and one for the torque converter lock-up clutch, on powershift gearboxes.
- Engine control system communication line, to regulate engine torque during speed shifts when critical to the life of clutches.

- Instrument panel communication line to show selected gear and possible malfunctions.
- Selector interlock to inhibit N/D or N/R or P/R shifts if the brake pedal is not depressed; the vehicle could in fact move if not braked when the torque converter is applied.

Both input and output list will surely be supplemented as control system functions increase in number and complexity. Many inputs used by the gearbox control system could be made available for other vehicle controls; in this case signals could be gathered by a serial communication bus.

15.6.1 Speed selection for minimum consumption

The first function to be performed by control systems is speed selection. The term has no single definition because the speeds available in certain engine operation regimes can be numerous.

Consider, for example, the diagram in Fig. 15.28, where available power curves at different gearbox speeds and the power necessary for motion at different slopes are shown as a function of car speed; at zero slope, at speeds between 15 and 20 km/h, the engine can work in first and second speed; between 40 and 60 km/h all speeds are usable.

Speed selection must also take into account the requested acceleration; to deliver, for example, 50 kW at 50 km/h the first speed is necessary, but to deliver 40 kW only first and second can be utilized. To define gearbox speed in an unambiguous way another condition must be established – minimum fuel consumption.

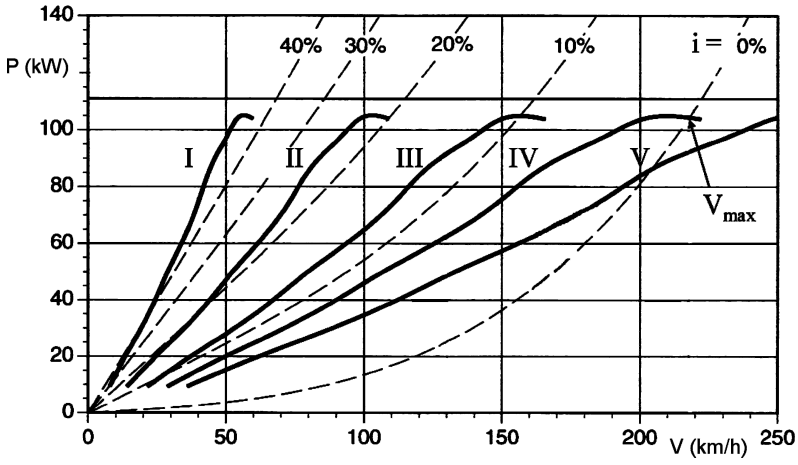


FIGURE 15.28. Diagram representing available power at different gearbox speeds to accelerate the vehicle on an assigned slope at a given vehicle speed; available power is the distance between one of the thick lines and the selected slope curve.

If we recall the specific fuel consumption map in Fig. 15.20, we see that the minimum consumption curve is slightly lower than the maximum power curve. At minimum engine speed, assumed to be constant, the minimum consumption curve is a vertical segment and the engine must be regulated by load alone.

If we imagine drawing the map of Fig. 15.28 on the diagram of Fig. 15.20 and if we consider the already defined transmission ratios, we can say that first speed is suitable for minimum consumption within the working area limited:

- On the left by maximum power available in first and by minimum feasible engine speed in first
- On the right, by minimum feasible speed in second, by maximum available power in second and by maximum engine speed in first

The same conclusions can be drawn for all speeds up to the fifth, where one of the boundaries is given by the power absorbed by driving resistance.

Desired power is not an actual input available to the control system; moreover, drivers are not used to thinking in terms of power, but act on the accelerator pedal by increasing or decreasing its stroke after comparison of actual performance with that desired.

In fact, when drivers think about accelerating the vehicle, they depress the pedal until the desired feeling is obtained; if maximum acceleration is sought they push the pedal all the way down.

Let us consider, for example, a petrol engine with throttle valve and mechanically linked accelerator pedal; the torque delivered by the engine does not move in fashion with the engine speed, as we can observe in Fig. 15.29 on different diagrams at different throttle angles α. The same torque can be offered at different speeds and different throttle angles.

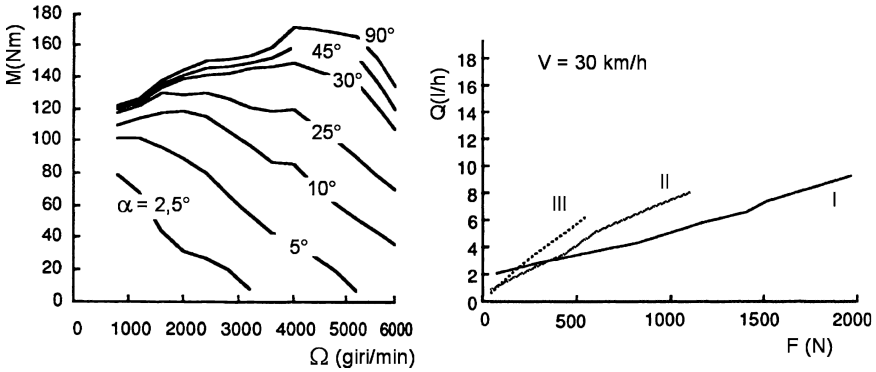


FIGURE 15.29. Map of the torque delivered at different throttle angles (on the left); with simple elaborations it is possible to calculate the fuel rate curves necessary to obtain a given traction. Starting from the intersection of these curves it is possible to obtain throttle angles at a given speed where shifting speed is convenient.

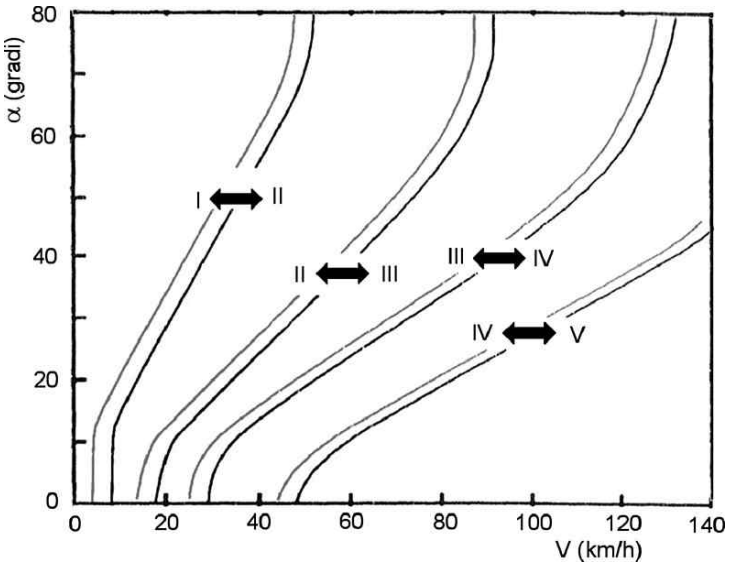


FIGURE 15.30. Qualitative diagram of optimum shift curves for minimum consumption; the black curves are for upshifts, the grey ones for downshifts.

If we imagine elaborating the fuel consumption map so as to build up, for each vehicle speed (in Fig. 15.29, on the right, the example for 30 km/h is shown), the traction diagram as a function of fuel rate, we can draw a set of curves for each of the available transmission ratios at that vehicle speed.

The intersections between traction curves show shift points to keep the consumption at a minimum.

Shift point throttle angle can be easily calculated.

From the fuel consumption map and the throttle valve map, by repeating this process for a sufficient number of vehicle speeds, it is possible to draw the curves of Fig. 15.30, which limit the optimum utilization regions of each gearbox speed on the throttle valve versus the vehicle speed plane; both magnitudes are known to the control system.

To avoid, finally, any oscillation between two neighboring gearbox speeds, the obtained curves are now interpreted as upshift curves; other curves are moved to lower speeds for downshifts (in grey). Within the area between these curves the actual speed is maintained.

15.6.2 Speed selection for comfort

If we now consider the problem of defining the most suitable shift speed for comfort, we must request that traction at a given gearbox speed be the same as for the next speed, at the same accelerator pedal position; as a matter of fact, with an automatic gearbox, the driver does not anticipate shifting time and does not usually move the accelerator pedal during speed shifts.

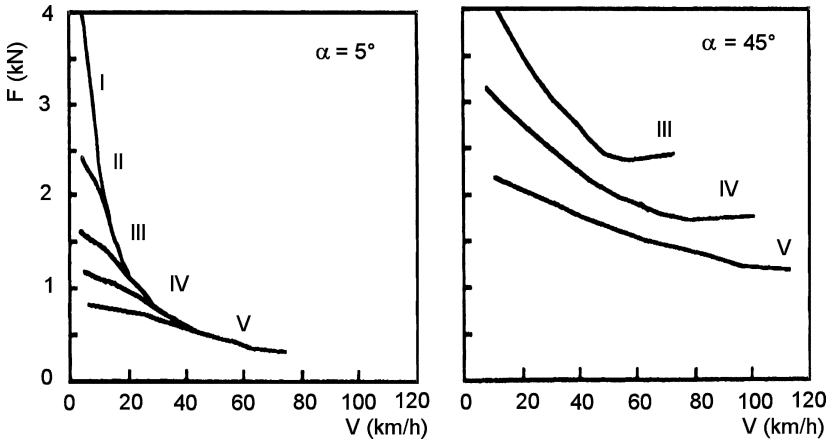


FIGURE 15.31. Family of traction curves at the same throttle angle, for two different values of this last parameter. At small angle values the curves show an intersection; at higher angle values the intersection may not exist.

Let us again consider a petrol engine; the case of a diesel engine is not substantially different. We now transform the throttle angle map into a family of traction curves at the same throttle angle (we produce, in other words, the traction map as a function of speed, having throttle angle as a parameter): Two different situations arise that are explained in the following examples.

On the left side of Fig. 15.31 we see the family of curves for a small throttle angle of 5 deg (usually the wide open position is at 90 deg); the curves show an intersection that represents the most suitable point for shifting.

In the case of an opening of 45 deg (right side), the curves have no intersection; in this situation the optimum shift point should be set where the curves are minimally spaced; at wide open throttle the shift point is positioned at the maximum engine speed.

If we group the optimum shift points on a throttle angle diagram as a function of vehicle speed, we obtain a family of curves similar to that of the previous paragraph.

The family of minimum consumption curves is usually shifted to lower speeds as compared with that of maximum comfort, except at the maximum power point (wide open throttle) where the two families are coincident.

15.6.3 Definition of a compromise choice

The choice of engaged speed on a gearbox can be made according to different approaches.

A first simple approach is to compromise between the two criteria, defining a curve family between them. But at very low vehicle speed the jerk produced by lowering the shift speed may be disappointing, because the acceleration is

modest; in addition, fuel consumption may suffer by some percentage if leaving the optimum condition.

The most widely used approach was, in the past, to assign the responsibility of choice to the driver, by introducing a program selector which determines the two alternatives. The economy option can be selected, for instance, in city traffic or slow suburban traffic, while the comfort option may be chosen elsewhere.

This approach is questionable because it introduces an additional control with no immediate feed-back, which can be easily left in the wrong position.

On more modern control systems, algorithms are implemented that recognize the driver's intention and driving environment by quick statistical elaborations on operating parameters already known to the control system.

15.6.4 Speed choice in real driving conditions

A fixed shift speed curve at a given vehicle speed and throttle angle does not represent an acceptable choice in conditions other than driving on a flat road with no traffic.

Let us assume that the vehicle is moving on a flat road at a high gear ratio. The operating conditions are represented in Fig. 15.32 by the point 1 of the diagram; the dotted curves represent the traction force F_x at the wheel, as a function of the throttle angle α.

Let us also assume that the car is facing a sudden slope, raising the traction to $F_{x,0}$, as shown on the diagram; the driver will feel the car slowing down and will react increasing throttle angle up to the corresponding value of point 2, where a downshift is begun.

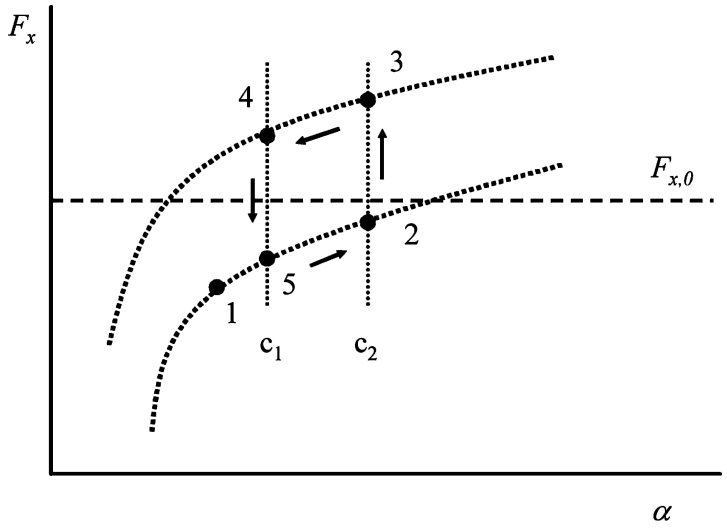


FIGURE 15.32. Typical gearbox speed oscillation cycle occurring when desired traction value is between the feasible values of two neighboring speeds at the same throttle angle.

When the shift has been completed the driver will most likely release the accelerator pedal, in the attempt to keep the desired car velocity, starting a new upshift that again reduces the traction force.

A series of unrequested speed shifts will occur (cycle 2, 3, 4, 5), well known to those who have driven a car on hilly roads at an imposed low speed.

The solution of the problem should be to temporarily move the shifting line to a lower throttle angle so as to reach the desired traction value on the lower speed, as is accomplished on a manual gearbox.

A similar situation occurs when on an uphill road a sudden short descent occurs or when a curve is encountered on a flat road. In the first case the accelerator pedal is released and the next speed may be shifted to, with a following upshift when the uphill road restarts. With a manual gearbox the speed would not be changed. On the unexpected curve, the same situation could occur.

The above inconvenience can be avoided when setting the selector manually so as to maintain the low gear.

The solution is not satisfactory because it implies additional operations and because drivers may forget and leave the selector in the wrong position, particularly if a high gear is chosen.

As an alternative, drivers expert in the use of an automatic gearbox increase their distance from preceding cars so as to avoid any action on the accelerator pedal; the power reserve for acceleration will decrease, making overtaking more difficult and sometimes reducing average speed.

Drivers using a manual gearbox can see the oncoming descent or curve and decide whether to maintain or change speed; this possibility is not available to automatic gearbox control systems. The spread of satellite navigation systems could, in the future, make this option a possibility.

A simpler and effective system of solving this problem comes from the observation that on a hilly or bending road the accelerator pedal is moved much more frequently and with ampler displacements, if the driver desires to drive quickly.

The presence and the value of a slope may be estimated by the average value and variance of throttle angle. By comparing calculated values with a reference value that can be achieved empirically, it is possible to decide to lower the downshift points.

Another road slope indicator may be engine torque; it can be estimated by control systems by a statistical analysis of injection time and other engine parameters.

Sport driving behavior can be detected by the throttle angle first derivative.

If after accelerator depression, the vehicle response is not adequate (the existing gearbox speed has been maintained) a second actuation step on the pedal will follow; this event can be an indicator of the driver's lack of satisfaction.

This situation may suggest adopting a higher upshift curve; in the same way the persistence of low throttle angle speeds may suggest returning to lower speed shifts more suitable to fuel consumption.

Another particular situation may occur when the accelerator pedal is released to obtain weak slow down or slow down with the use of brakes. In the first case it is better to leave the actual gearbox speed or to select a higher one to lower fuel consumption; in the second case a downshift may be necessary to enhance the engine braking force.

The first behavior is already implemented in the shifting curves discussed in the previous paragraphs.

The second behavior can be implemented by considering brake pressure and the throttle angle first derivative; a desire for stronger slow down is characterized by higher accelerator release speeds.

15.6.5 Brake and clutch actuation

A last problem to be tackled with the control system is clutch or brake actuation, involved during speed shifts and torque converter lock-up clutch actuation.

In previous paragraphs we have defined shift conditions for maximizing comfort or minimizing consumption; the above conditions still do not consider the shifting transient generated by clutches.

It is necessary that clutch disengagement and engagement, involved in speed shifts, follow certain criteria to avoid a disturbing jerk during the manoeuvre.

Let us consider the example of Fig. 15.33. An epicycloidal gear train is represented with four different speeds; we consider only the first to second speed shift in order to understand what is going on during this period.

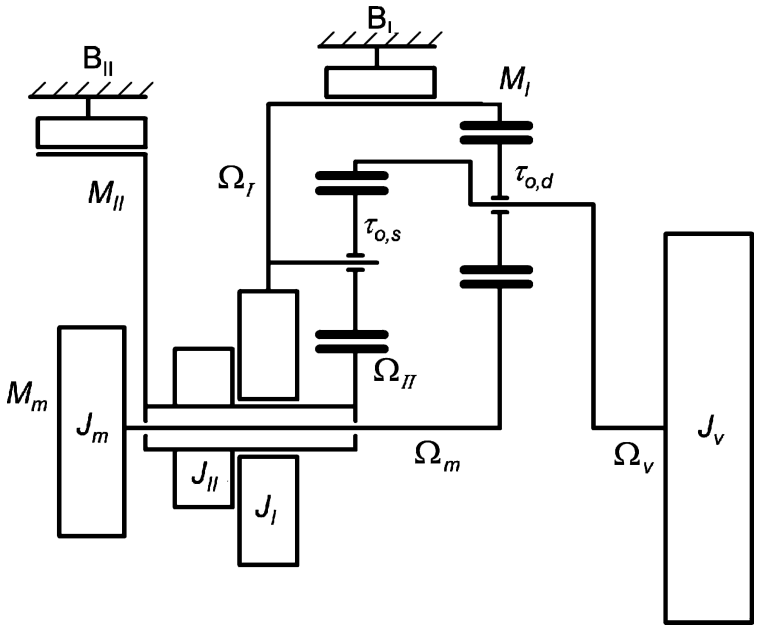


FIGURE 15.33. Epicycloidal gear train scheme suitable for four speeds; the rotating masses involved in the shift from first to second speed are shown.

In first gear the band brake B_I is closed.

Let us consider the ordinary transmission ratios of the left and right trains, called $\tau_{o,d}$ and $\tau_{o,s}$ respectively.

Subscripts m and v show elements turning at engine speed and at vehicle transmission shaft speed; for the sake of simplicity the torque converter is ignored.

In first gear the band brake B_{II} drum that will be actuated to obtain the second speed is turning at a speed given by:

$$\Omega_{II} = -\Omega_v \tau_{o,s} \ . \tag{15.16}$$

In second gear the brake B_{II} is closed while the drum of the band brake B_I will be left free to turn.

The shift procedure from first to second will start working on brake B_{II}; the engine, gearbox and vehicle system will be slowed down by braking moment application but will receive a positive contribution from the partial transformation of kinetic energy of brake B_{II}'s rotating parts. These are modelled with flywheel J_{II} while the engine is modelled with flywheel J_m; in the same way a positive contribution will result from the torque reduction of brake B_I, and a negative contribution will result from the absorbed work to accelerate flywheel J_I to the speed:

$$\Omega_I = \Omega_v + \frac{1}{\tau_{o,s}(\Omega_m \ - \ \Omega_v)} \ . \tag{15.17}$$

Input magnitudes for a high quality upshift are therefore many; in Fig. 15.34 is reported as an example a typical diagram of most relevant magnitudes as a function of time, for the first to second upshift at wide open throttle. Engine speed increases before and after the shift, if the supplied power is larger than resistance; engine speed will decrease as a consequence of the transmission ratio decrease.

Gearbox output torque will be reduced as a consequence of the upshift; an ideal solution, with the aim of reducing jerk, will be to create a torque diagram with a straight line joining the torque values before and after the shift point; the real value is represented by the oscillating line.

In the two hydraulic circuits for clutch or brake actuation there will be:

- A decreasing pressure versus time on elements locked on first speed.
- An increasing pressure versus time on the element to be locked for the second speed; the two diagrams should overlap precisely to minimize traction variation.

The curve of the rise and fall of pressure is assigned to proportional control valves that can obtain a traction curve of continuous shape.

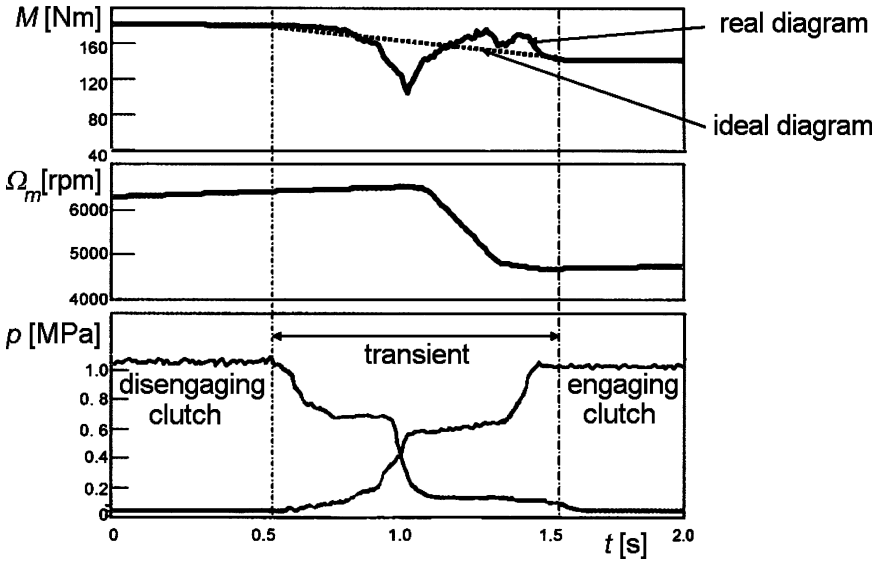


FIGURE 15.34. Diagram of the most important magnitudes involved in an upshift versus time t; from the top, gearbox output torque M (real and ideal values), engine speed Ω_m and oil pressure p on brakes are represented.

Optimum actuation lows for brakes and clutches must be determined for each speed shift as a function of torque and speed. Mapping actuation functions can be too burdensome in terms of experimental activity and control system memory; mathematical models may be more convenient.

While designing these models it must be taken into account that gearbox output torque is determined not only by input torque, but also by the inertia of gearbox rotating elements, which cannot be neglected, particularly for epicycloidal gear trains.

The best result, in terms of jerk reduction, can be obtained when clutch (or brake) pressure can be regulated proportionally and independently; acceptable results, such as those shown in the figure, can be obtained with a continuous regulation of the closing element and a pulsed discharge of the opening element.

Also in case of non-epicycloidal gear train gearboxes (fixed rotation axis and dual clutch) the inertia contribution is not negligible.

Let us consider a two speed gearbox, as represented by the scheme in Fig. 15.35; M_I and M_{II} are the instantaneous slip torques (function of time) of clutches B_I and B_{II} involved in the transmission ratio variation from τ_I to τ_{II}. The apparent vehicle rotating mass is represented by J_v while J_m is the equivalent inertia of the engine, M_m is the engine torque and M'_m the driving torque at the gearbox output shaft.

We assume that during shifts there is no important change in resistant torque. Engine torque can sometimes be considered as being constant too; in the

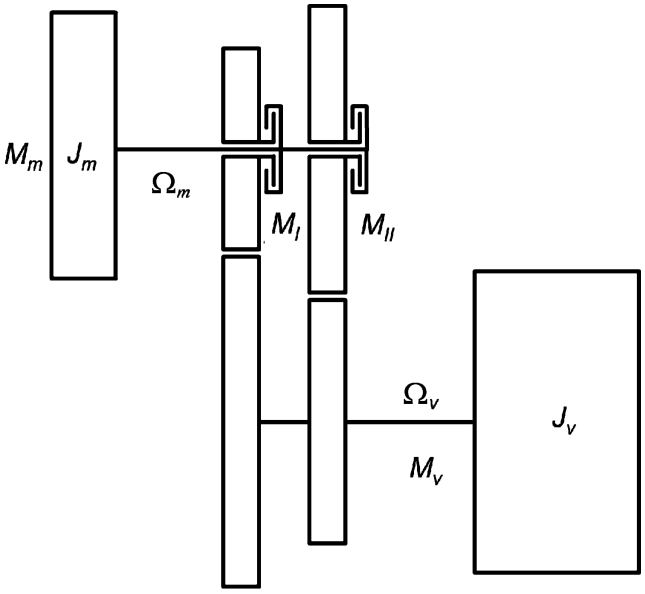


FIGURE 15.35. Scheme for the study of speed shift transient in a powershift gearbox with fixed rotation axis.

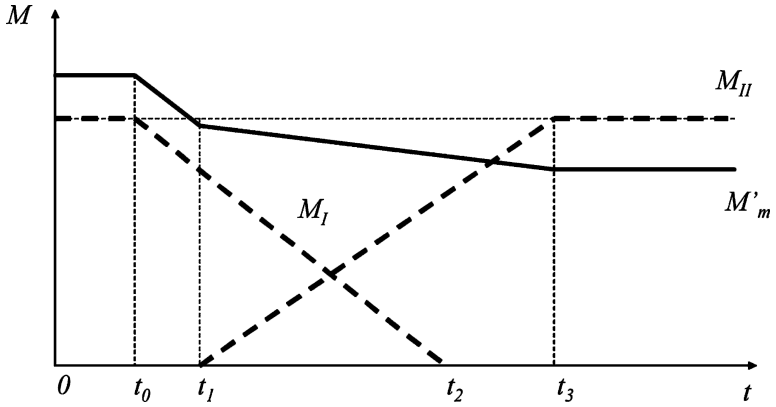


FIGURE 15.36. Schematic diagram of torque on engagement clutches of a countershaft powershift gearbox during an upshift.

most general case, M_m is a known function of Ω_m, if the throttle valve angle remains unchanged during the manoeuvre.

We assume finally that the clutch actuation forces are linear in time.

In Fig. 15.36 we can draw a schematic diagram of the phenomenon; we can write:

$$\begin{cases} J_m \frac{d\Omega_m}{dt} = M_m(\Omega_m) - M_I(t) - M_{II}(t) \\ J_v \frac{d\Omega_v}{dt} + M_v = \tau_I M_I(t) + \tau_{II} M_{II}(t) = M'_m \end{cases} . \tag{15.18}$$

The derivative of the second equation is the jerk; it could never be zero because the M'_m torque will change, if only because of the change of the transmission ratio. We can understand that the minimum jerk is achieved if M_I and M_{II} are linear in time and when t_1 and t_0 are equal.

Piston filling and emptying times are determined not only by their volumes and by orifice dimensions, but also by:

- Oil circuit pressure, which is determined by pump speed (rotating with the torque converter pump) and by oil viscosity, a function of temperature; pressure is in fact determined by internal oil spillage in the pump.
- Clutches rotational speed, because, very often, the centrifugal pressure field is not negligible.
- Friction coefficient on clutch working faces, dependent on wear and on temperature.
- Oil viscosity and temperature that condition filling and emptying time at a given pressure.
- Dimension tolerances of the gearbox.

For these reasons the development of a reliable mathematical model can be particularly difficult. It is a good practice to use a first approximation mathematical model, integrated by feed-back parameters; these can be considered from time to time to compensate for the tolerance or, more frequently, for temperature effects.

The ideal feed-back parameter should be output torque, but at this time viable torque sensors do not exist; torque can, however, be estimated by accelerations, once inertia masses are known.

A further parameter to be used to minimize acceleration variations is controlling engine torque, through spark advance or, if possible, directly on the throttle.

It is also important to eliminate some of the variables of the system, by regulating oil pressure at a constant value, eliminating some of the effects of temperature and viscosity.

A few words must be written on lock-up clutches; to their actuator apply all considerations made for clutches and brakes.

The lock-up clutch must be engaged when the converter slip ratio goes to low values, indicating that torque multiplication is no longer necessary.

The clutch engagement also eliminates damping capacity, which is quite useful particularly for low speeds where the vehicle equivalent rotating mass is high.

On old gearboxes the lock-up clutch was only actuated in top gear. More attention has been given on more recent models to fuel consumption, causing the use lock-up clutch to be used at all gearbox speeds; the benefit on consumption is of several percent.

In this case the actuation pressure must be carefully controlled as a function of torque not only at the engagement but also during normal operation, to allow a small slip when torsional vibrations may occur.

A damper similar to that of a conventional start-up clutch is in this case recommended.

16
DESIGN AND TESTING

In this chapter we will discuss design rules and procedures of those transmission components not considered in previous chapters, and test methods that can be applied for validation and qualification of the entire transmission as a system.

Bear in mind that the objective of this chapter is to complement the knowledge gained in engineering courses on machine design and calculation; we refer back to these for the fundamental know-how.

16.1 TRANSMISSION MISSION

We will describe in this section the operating condition of the transmission during the vehicle's life, to understand qualitatively which loads are applied and how long they must be withstood without damage.

The transmission plays a significant role in the operation of a vehicle and must therefore have an average useful life at least as long as that of the vehicle itself; a suggested maintenance program must obviously be followed.

Transmissions must preserve not only their structural integrity but also their functional characteristics, of which ease of use, as we have described in our discussion of shifting mechanisms and synchronizers, mechanical efficiency and generated noise are the most relevant.

Transmission life, as is true of other vehicle systems, can only be described in statistical rather than deterministic terms, because loads deriving from road type and driving style have a statistical nature.

G. Genta and L. Morello, *The Automotive Chassis, Volume 1: Components Design*, Mechanical Engineering Series,

© Springer Science+Business Media B.V. 2009

Therefore the endurance specifications, normally adopted by manufacturers, are assigned statistically through the magnitude B_{10} which corresponds to the endurance not attained by the 10% of produced transmissions.

The *endurance* is the maximum distance traveled according to the foreseen mission, without any major damage.

If we take into account recent product evolution, we should not limit the major damage category to failures that interrupt vehicle motion, but should also include all those that are usually claimed by customers at the service shop, such as:

- Noise increase
- Increase in shifting loads
- Lubricant spills, etc.

As reference values for B_{10}, we can assume:

- Automobile transmissions >150,000 km
- Construction trucks >300,000 km
- Urban buses >400,000 km
- Long haul trucks >800,000 km

The final result to be obtained is clear enough; it must be said, however, that the quantitative specification of transmission life is made difficult by working conditions which are always changing and sometimes are unforeseeable.

If we limit ourselves to automobiles, we can say that in Europe that life is spread over these four typical environments:

- 40–70% on motorways
- 15–30% on suburban roads
- 15–20% on urban roads
- 3–10% on mountain roads

Each manufacturer independently promoted test campaigns addressed at gathering reliable data on the usage of cars; the results of these tests are an important part of company know-how and are seldom published.

These data acquisitions must be made for different groups of homogeneous customers using different types of cars and must be repeated at different times, to take into account changes in living habits, road networks, traffic and preferred types of cars.

It is usual to employ data gathered in the fleets of special customers, such as taxis, company cars, etc., which allow data measurement on their cars (through inboard digital recorders), in deciding the most favorable selling price.

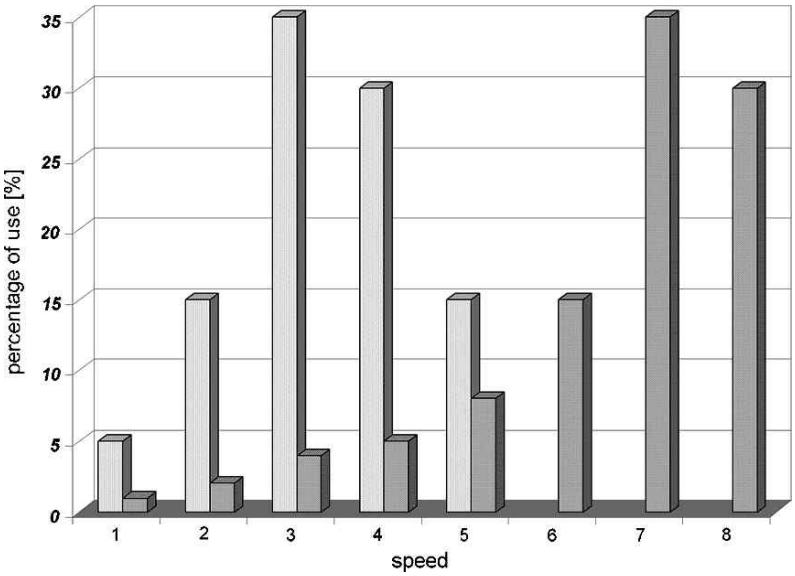


FIGURE 16.1. Qualitative diagram of different speeds, in the average life of a car (5 bars histogram) and of a medium duty industrial vehicle (8 bars histogram).

A typical data acquisition includes time histories of car speed, gear shift position, clutch position, accelerator position and road slope; these are statistically synthesized to obtain easier to use information. From these, many other data are derived, such as those related to transmissions. Transmission operation is identified by input torque, engaged speed and clutch pedal position.

The time history of these parameters can be used as a fixed reference for design calculations and validation of experimental endurance.

As an example, in Fig. 16.1 results are reported for a car and a medium duty truck, showing the percentage of use of available gearbox speeds.

The input torque changes strongly over time and can also be negative; the service life of all structural transmission components is therefore limited by fatigue.

Gears and other rotary elements introduce additional load variations over time.

Table 16.1 shows a qualitative description of the transmission life, after applying some simplifications.

As a first assumption, to stay on the safe side, only mechanisms stressed by I and II speed are designed for a limited life: Mechanisms stressed by III, IV and V speed are designed for unlimited life; because all components are basically made of steel a life of 100 million cycles is assumed.

As far as reverse speed is concerned, a life as long as one quarter of the I speed life is usually assumed.

TABLE 16.1. Qualitative description of the average use condition of a car transmission. The life spent in I and II speeds is measured in (km), that spent in III, IV and V speeds is measured in millions of cycles.

Speed	Max torque	Max power	Braking	Total
I	5,600	1,600	800	8,000
Shifts N/I				260,000
II	21,000	6,000	3,000	30,000
Shifts I/II				180,000
III	70	20	10	100
Shifts II/III				100,000
IV	70	20	10	100
Shifts III/IV				20,000
V	70	20	10	100
Shifts IV/V				20,000

The endurance requested for the I and II speeds are simply derived by multiplying the target life of 200,000 km by the percentages shown in Fig. 16.1.

Load conditions are simplified and made much more severe by assuming that the car is working:

- 70% of the time, at maximum torque and at the corresponding engine speed
- 20% of the time at maximum power and at the corresponding engine speed
- For the remaining 10% of the time with the engine braking the car with a torque as high as half of the maximum torque and at maximum engine speed

Gear shifts are estimated by assuming that the gearbox is used during its entire life of 200,000 km as in the European fuel consumption and emission cycle, and is, therefore, driven 20,000 times. As a consequence of the assumption made in the third point, a number of downshifts is added equal to 10% of the upshifts.

These load conditions can be used for designing gears, shafts, bearings, housings and seals, while the number of up and downshifts can be used to design synchronizers, shifting mechanisms and clutches.

This table should be improved with an accurate statistical study for the application under scrutiny, but it can be considered as a starting point for a preliminary design of a car transmission.

Input parameters acquired during real driving conditions are useful for estimating the average life of components; as a good praxis, we recommend also considering some exceptional conditions that may occur when transmissions are used in an incorrect but possible way; in these conditions as well, catastrophic failures are not accepted.

These test procedures are part of the manufacturers' know-how; they are performed by driving the car according to heavy driving schedules that condense the car's useful life into a short time. A typical test of this kind consists, for example, in releasing the clutch pedal suddenly, as may happen while wearing muddy and slippery shoes.

Transmission components are classified by the design process, according to three different categories:

- A category, including components that are critical to correct transmission operation and can be calculated following reliable and validated procedures (they are: shafts, gears, bearings).
- B category, including components that are critical to correct transmission operation, but cannot be calculated following reliable procedures (they are: synchronizers, lubricants, seals).
- C category, including non-critical components, because previous experience does not record failures (they are: venting valves, housings, etc.).

B category components are designed to imitate previous successful examples; the prototypes of these components must be tested extensively on bench and vehicle to assure their reliability before building the first transmission prototype.

It is therefore important for an effective management of development plans that all components accomplish the experimental validation process with the same level of confidence in success.

All components, including those which benefit from a reliable design process, will be tested in a real transmission on a bench and finally on a real vehicle, as soon as they have reached their final design, to demonstrate the achieved reliability.

Reliability is the probability of surviving a certain mission without failures.

The demonstration of reliability must be made on a sufficiently wide sample of prototypes; these must be built with their critical characteristics on the border of their zone of tolerance in a real production process.

The design of these tests, aimed at guaranteeing a sufficient confidence level, can be studied in more specific handbooks.

16.2 GEARS

16.2.1 Endurance

Gear life is limited by four different kinds of tooth failure:

- Bending
- Pitting

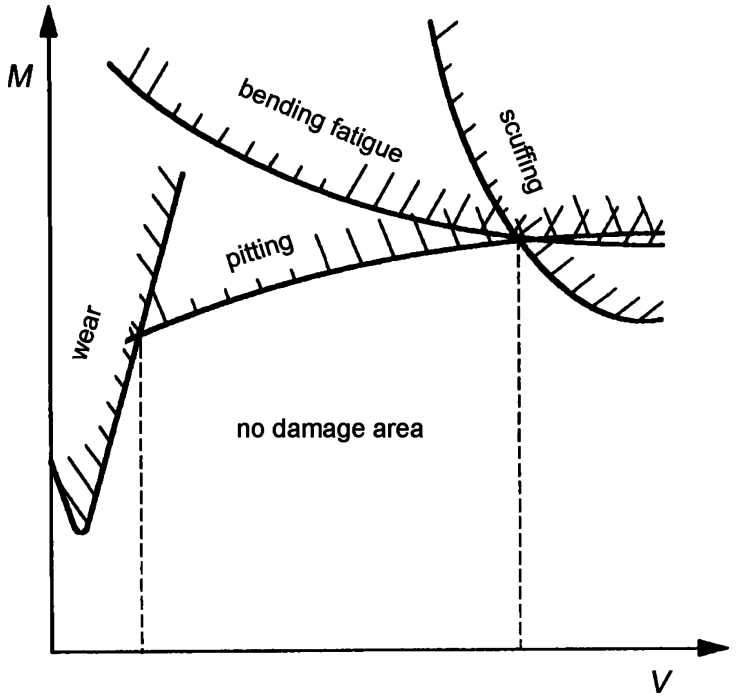


FIGURE 16.2. Qualitative diagram of the transmitted torque M as a function of circumferential speed V showing the typical shape of operation boundary curves for bending fatigue, pitting, scuffing and wear; the area below the curves envelope corresponds to operation without danger of damage. In consideration of the most likely speed values, pitting is the most important limiting condition.

- Scuffing
- Wear

Figure 16.2 shows the qualitative extension of survival fields for these phenomena, on a diagram of transmitted torque versus circumferential speed.

The following issues are relevant for determining gear life:

- Working conditions (transmitted torque, circumferential speed, oil temperature)
- Tooth material
- Tooth geometry
- Surface treatment
- Lubricant formulation

FIGURE 16.3. Appearance of a tooth flank damaged by bending fatigue.

For the calculation of spur gear wheels, the standard procedure ISO 6336 is used by almost all car manufacturers.

Tooth bending failures occur when a substantial part of the tooth is removed as a result of the failure; it is useful to separate failures due to overload from those due to fatigue.

Failures of the second kind have the aspect shown in Fig. 16.3 and clearly show a point where the failure started (a surface defect, a notch or a non-catastrophic overload failure).

A set of lines which are almost concentric to the starting point show the propagation direction of the failure. Failure surfaces are polished by the continuous shock of the two parts, due to load pulsation.

A part in terminal failure can be seen, characterized by a very irregular surface with crystal aspect, due to a sudden rupture when the resistant section has decreased too much.

Overload failures have the appearance of the terminal part of fatigue failures.

The stress variation at the origin of the fatigue phenomenon is due to the following causes:

- Low frequency torque variation; torque can also change direction
- Cyclic teeth mesh and variation in the number of teeth in contact, at medium frequency
- Engine torque harmonic content, at high frequency

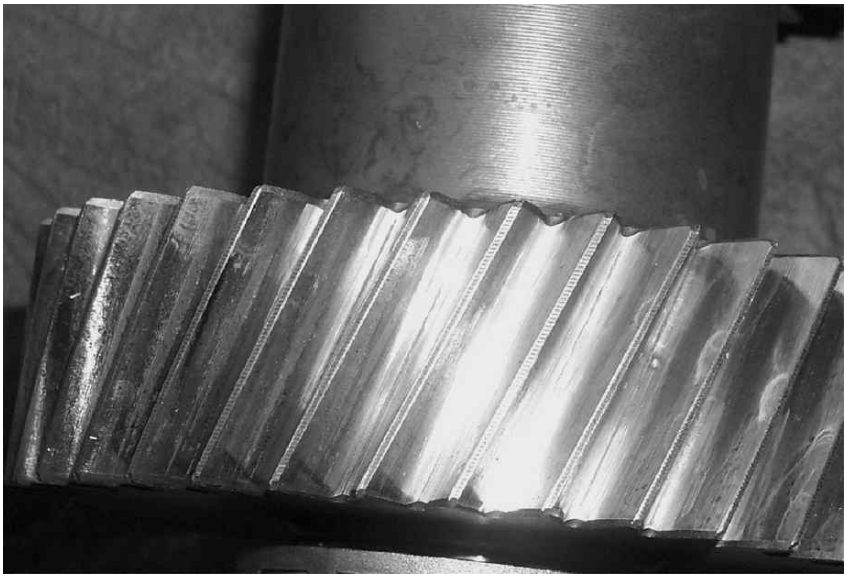


FIGURE 16.4. Aspect of a tooth flank damaged by pitting.

By their nature, idlers work under alternate loads.

Pitting damage is shown by the appearance of growing craters which can merge. They are caused by failures under the tooth surface, the result of too high a hertzian contact pressure.

This failure can be characterized by very small craters, which give the surface a velvety aspect, as seen in Fig. 16.4, or by big and irregular craters such as those in Fig. 16.5; the difference in appearance is caused by the severity of the phenomenon in question.

It must be remembered that the pitting phenomenon occurs only on lubricated surfaces; lubricant characteristics therefore play a major role.

Low speed wear is characterized by abrasions on the tooth flanks, according to relative speed direction; it is seldom seen on automotive transmissions, except in cases of evident geometric errors or lack of lubricant.

The appearance of scuffed surfaces is quite similar to the previous; the damage can be caused by a rupture of lubricant film, due to high temperature or high pressure. It can cause metal on metal contact; its appearance is caused by micro welds and rips or by chemical action.

The quoted ISO norm includes calculation procedures addressed to verifying that none of the above phenomena occur; only scuffing is not considered by calculations.

In the diagram of Fig. 16.2 relative to automotive transmissions, the safe tooth operation area is essentially limited by pitting.

Because of applied load time variation, we must consider the phenomenon of fatigue at the design stage.

FIGURE 16.5. Aspect of a tooth flank damaged by heavy pitting.

Therefore the so-called Wöhler curves must be considered; one of these curves is shown in Fig. 16.6.

The two lines (inclined and horizontal) on the right of the figure represent, on a bilogarithmic scale, stress amplitude (on the vertical axis) as a function of cycle number leading to failure; there is, as known, a horizontal limit, the fatigue limit, below which, according to a simplified theory, failure will never occur.

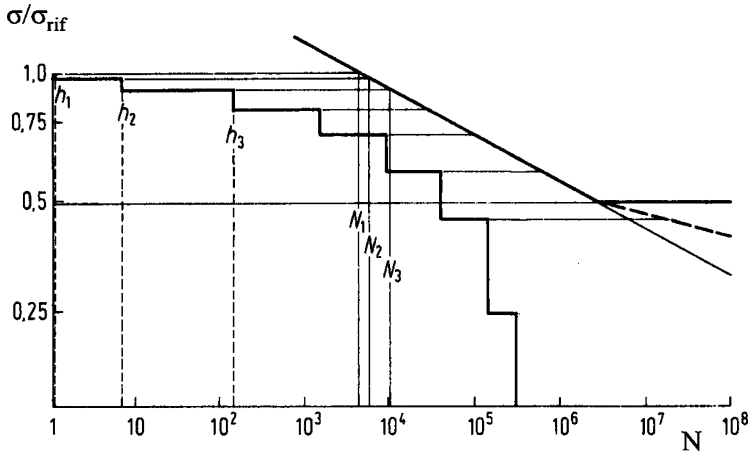


FIGURE 16.6. Comparison of a Wöhler (bilogarithmic representation) curve with an assigned load profile (staircase line). Stress is divided by a reference value; N_1, N_2, N_3 represent the cycle number during which relative loads h_1, h_2, h_3 are applied.

It is good praxis to consider not only material probe average failure curves, but also B_{10}, B_{50}, B_{90} curves, if not only average life, but reliability as well has to be predicted.

In order to compare working stress with the Wöhler limit, it is useful to transform, as a first step, the torque time history into stress time history.

It is possible, at this time, to count the cycle number at the same stress amplitude (h in the diagram) so as to define a histogram in which each stress amplitude class is represented with the corresponding cycle number (N in the diagram); this histogram is the so-called *load profile*.

A load profile, reordered in decreasing load amplitude classes, is represented by the staircase line below the Wöhler curve.

At this time the comparison of actual with the allowed stress is possible.

The rule that is normally applied in this comparison was developed by Miner; according to this rule, each group of loads of the same amplitude produces a partial damage to the material, equal to the ratio between actual number of cycles and damage-free cycles, on the Wöhler curve.

The same rule says that failure occurs when the sum of partial damages (each of them lower than one, by definition) is equal to one.

Miner's rule implies no damage for load amplitudes below the fatigue limit; this hypothesis should be put under discussion, because if partial damage has already occurred, cracks have already been started; it is also possible that stress levels below the fatigue limit, if referred to the undamaged component, are causing the damage to proceed.

To take this fact into account, empirical corrections are used; according to these corrections the horizontal fatigue limit line is changed to an inclined line,

whose slope is defined by the bisector line (on the bilogarithmic representation) between the inclined and horizontal lines. This line is reported on the diagram with dots.

While choosing in the literature the appropriated Wöhler curve for gear material, it is important to take into account the actual heat treatment (case hardening or nitriding, this last applied usually to epicycloidal gears) and the finishing of surfaces in contact, which can be ground, shaved or honed.

A common characteristic of countershaft automotive gearboxes is to adopt the same centre distance for all gears. The centre distance of a gearbox is decided by the most stressed wheel couple, usually that of first and reverse speeds.

The smaller stress for other speeds will be taken into account with a smaller face width; the ratio between face width and diameter is conditioned by the space available to install the transmission: for this reason this ratio is not very dissimilar in different gearboxes (between 0.3 and 0.6).

The centre distance is therefore an engineering parameter indicating design overall performance; it may be used to compare gearboxes of same torque capacity.

16.2.2 Noise

Many different noise qualities produced by the gearbox are judged unpleasant and can draw the driver's attention; they can be classified precisely according to different categories:

- Gearing *whistle*
- Gearing *rattle*
- Speed shift *grating*
- Bearing *whine*

Whistle can be caused by three different factors:

- Pitch errors in contacting teeth and pitch variations due to deformation of contacting teeth.
- Tooth bending stiffness variation along the contact line; under this aspect helical spur gears show better behavior than straight gears.
- Undulations of contacting teeth surfaces.

Rattle noise is particularly disagreeable and is produced both by idle gear wheels in constant mesh and by sleeves and synchronizers hubs; Fig. 16.7 shows, on the left, the case of a single stage gearbox and on the right the case of a double stage countershaft gearbox.

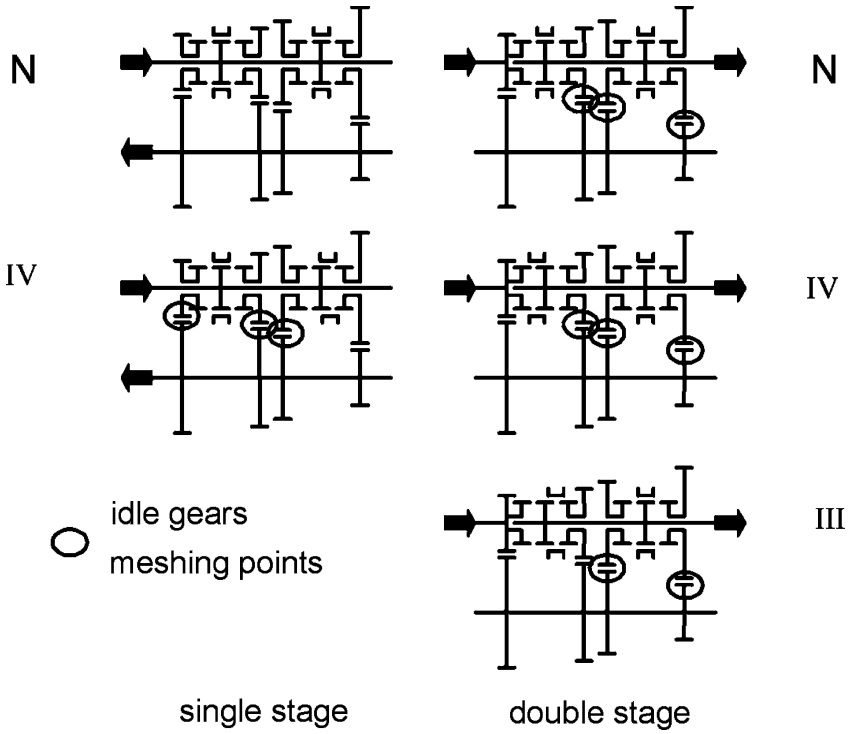


FIGURE 16.7. Schemes of a single stage and a double stage countershaft gearboxes; circles mark meshing points of rotating idle gears, responsible for generation of rattle noise.

The engaged gear wheels are shown by the position of their sleeves; circles mark the meshing points of rotating idle gears. At the top of the figure what happens on a gearbox in idle speed is shown.

Constant gear puts the countershaft into rotation. Because engine speed is not constant, the countershaft will have torsional oscillations that cause the teeth flanks to hit in the two angular directions left free by clearance.

A similar phenomenon may occur between sleeve spline and synchronizer hub and may be perceptible on idle couples, in direct drive or at other speeds.

Single stage gearboxes, in the left figure, present the same kind of phenomena, except at idle speed and for wheel couples that have sleeves on the driving shaft; these will, in fact, be still.

Rattle noise at idle is similar to rattle noise in acceleration, but, due to the limited speed may produce a metallic sound; during acceleration the sound has many more harmonic frequencies and is similar to the noise of flowing fluids in pipes.

The probability of perceiving this noise reduces consistently at medium engine speed, because many harmonic frequencies exit the audible field and the noise is masked by other sources.

The influence of oil viscosity is notable, because when it is high, for example at low temperature, the vibration of the aforesaid parts is completely damped.

Grating noise is produced during speed shifts only, if the thrust force has overcome the synchronizer reaction to a premature completion of the shift manoeuvre.

This noise is caused by an inadequate synchronizer dimension or by a violent manoeuvre; it can be detected at slow speed and low ambient temperature, when the synchronizer cone does not succeed in removing oil film.

Bearing whine, finally, is a precursor of bearing failure only and is produced by rolling bodies when clearance with their rings is too high.

Noise reduction must be approached at a system level. Other subsystems at the source, such as engine, power train suspension, elastic parts of transmission line and half axles and, sometime, suspensions and tires contribute to transmission noise.

Airborne noise and noise structurally transmitted through the car body contribute to noise perception.

To think of reducing gearbox noise by working on the gearbox alone, could be misleading and cause excessive product cost; nevertheless, some good design rules should be taken into account.

As far as whistle is concerned, it is better to apply helical spur gears only, with a coverage ratio greater or equal to 2.5; it can be further increased by applying high contact ratio profiles, by increasing the ratio of teeth height and pitch.

A profile correction oriented to achieving a slim tooth tip could be also beneficial.

Entire transmission ratios should be avoided; the preferential wear of ever coupled gears can modulate the gearing sound.

It is better to use reasonably tight tolerances, such those included under IT5 and IT7 classes; tighter tolerance may be prescribed for the gear wheels in most frequent use, to contain product cost.

Teeth surface must be sufficiently smooth. Satisfactory results are attained by shaving and honing; grinding can be used in exceptional cases.

Rattle noise is bound, as we have seen, to circumferential clearance; it can be reduced by:

- Reducing clearance
- Reducing inertia rotary mass of passive parts
- Improving lubrication; the better washed the parts in contact, the higher the damping action
- More suitable tuning of the clutch damper, by which natural frequencies can be moved outside of a critical region
- Adoption of a double mass damping flywheel

Because noise is irradiated partly through the air and partly through the structure, it is good practice to stiffen gearbox housing panels as much as possible, by using suitable ribs and increasing the local stiffness of interface points with the power train suspension.

16.3 SHAFTS

Gearbox shafts are made with many changes in diameter, as they must allow many different parts to be fit; it may also happen that a tooth profile is cut directly onto them, because of their small diameter.

Figure 16.8 shows the details of a shaft for a front wheel drive gearbox with transverse engine. The many diameter variations must be smoothed with chamfers or rounding of edges, to reduce torsion and the concentration of bending stress, even reducing the resistant section.

Machine design manuals report the static form factor for the most widely used section transitions; these factors are coefficients by which the stress obtained by the application of the De Saint Venant theory must be multiplied, to obtain the actual stress.

Shaft calculation can be performed following the same procedure we have suggested for gear wheels, to take into account fatigue phenomena and changes in load amplitude.

While designing shafts, it is vitally important to introduce displacements calculation due to load application; as a matter of fact, displacements (both linear and angular) may change teeth working conditions, with a negative impact on noise and useful life.

From the study of displacement, some sound design rules arise that can be summarized as follows:

- To reduce the shaft span between bearing, by limiting gear wheel and synchronizer width as much as possible
- To install the wheels subject to highest loads as close to the bearings as possible
- To avoid too steep transitions of diameter, organizing components by increasing or decreasing diameters
- To avoid feather keys, preferring spline connections
- To smoothen diameter transitions and drilling, as shown in Fig. 16.8
- To use circlips at the ends of the shaft only

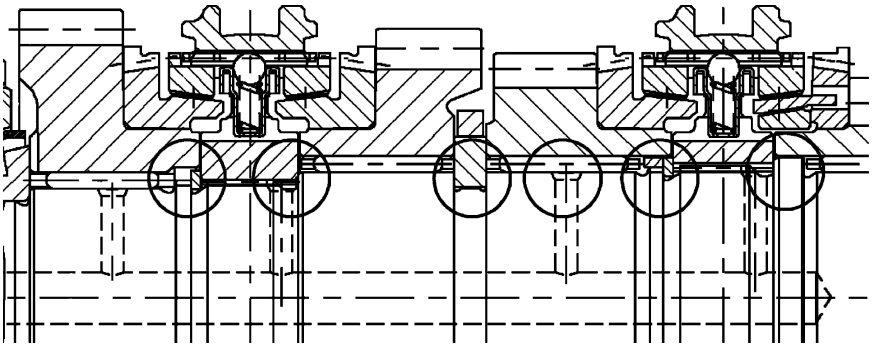


FIGURE 16.8. Some example of notching effect reduction on diameter transitions and transverse drills of a shaft of a single stage gearbox for front wheel drive (FIAT).

16.4 BEARINGS

Gearbox bearings are normally roller bearings; bush bearings are limited to lowly stressed idle gear wheels and to sliding bearings for internal shift mechanisms.

If possible, bearings are limited to two per shaft, avoiding hyperstatic mountings that are excessively sensitive to machining tolerances.

The problems to be considered when choosing a roller bearing are:

- Adequate component life
- Compliance to shaft angular displacements
- Compliance to differential thermal elongations of shafts (made of steel) and housings (made of aluminium or magnesium)
- Resistance to oil pollution caused by wear particles of components

The bearings applied to shaft ends are primarily of four different types:

- Deep groove ball bearing
- Four contact ball bearings
- Cylindrical roller bearings
- Tapered roller bearings

Deep groove ball bearings are widely applied because they can withstand both radial and axial loads. They are easily assembled on the shaft, require no position adjustment and are reasonably cheap; on the other hand they have the disadvantage of large dimensions and sensitivity to oil pollution.

In consideration of this point, self-lubricated sealed ball bearings are sometimes preferred, especially if gearbox splashed oil is abundantly available.

Four contact ball bearings are almost equivalent to the previous bearings, with the advantage of smaller dimensions at a slightly increased cost.

Cylindrical roller bearings have a high radial load capability but cannot withstand axial loads; they are usually coupled to ball bearings at the other end of the shaft. They have the disadvantage of a higher cost and inability to perform correctly in the event of substantial angular displacements.

Tapered roller bearings are applied more and more frequently because of their optimum radial and axial load capacity (see for reference Fig. 9.11). They have limited dimensions; their assembly on the gearbox is difficult, because the inner race with roller cage and outer race are separable components.

They require, for correct operation, an appropriate axial pre-load, which must be maintained at every work condition. The axial position of at least one the two bearing must be adjusted, to compensate for the length tolerance of shaft and housing; in addition, one of the two shoulders must be conveniently made, so as to be insensitive to thermal displacements. This result can be achieved by a spacer made with high thermal expansion material.

Bearings on idle gear wheels are primarily needle bearings; sometimes the races are the outer surface of the shaft and the inner surface of the gear wheel, to reduce radial dimensions.

16.5 LUBRICANTS

The correct lubricant specification is vitally important for gearbox endurance, considering the manifold functions it is expected to perform; the lubricant functions are the following:

- Reducing friction and wear of metallic and non-metallic (rotary and sliding seal) parts
- Distributing to colder areas the heat generated in hot ones, contributing to generated heat dissipation
- Building up lubricated hydrodynamic films
- Protecting components against corrosion
- Deterging residual particles produced by wear
- Performing all of the above functions for long periods and at any possible working temperature

Gearbox component lubrication is usually managed by splashing and spraying oil by rotating gears. It is therefore necessary to correctly exploit moving parts in order to provide and maintain oil for all couples that need lubrication.

This task is accomplished by designing oil channels into the housing in such a way as to pick up the oil projected by wheels and distribute it to less exposed

parts; transversal drilling on shafts may contribute to lubricating the bearings of idle gear wheels.

On heavy duty manual gearboxes and on automatic gearboxes pressure lubrication is used; a gear pump is moved by the input shaft and dedicated pipes and channels distribute the oil to the utilization points.

The study of lubricant distribution would be rather difficult if tackled by mathematical models; much simpler is to set up experimental analyses, using modified gearboxes with transparent windows.

When teeth flanks mesh together different phenomena can be identified:

- *Boundary lubrication*, when surfaces are in contact without the interposition of lubricant oil; their protection is granted solely by their nature; in this situation only lubricant additives can modify the chemical nature of surfaces and prevent micro welding (friction modifiers).

- *Mixed lubrication*, when a partial separation of surfaces takes place, through the effect of hydrodynamic forces generated by a lubricant.

- *Hydrodynamic lubrication*, when a completely separated lubricant film is available.

These three different situations take place according to the position of the contact point on the tooth flank and the peripheral speed; on the part of the tooth that is closer to the primitive circumference mixed lubrication will take place at low speed, while hydrodynamic lubrication will take place at high speed; on parts near to the first and last point of contact there will be boundary and mixed lubrication.

These facts should be considered when specifying a lubricant, as well as maximum temperature which can reach 90°C–100°C, in the bulk of the lubricant, and 150°C–160°C locally.

On present cars the expected lubricant life is as long as the life of the gearbox.

Lubricants that satisfy the above conditions are mineral oils mixed with synthetic oils; a suitable additives package must be provided to:

- Prevent corrosion and oxidation product build up

- Deterge and disperse pollutant particles

- Modify the chemical nature of surfaces in contact, to prevent micro welds in boundary lubrication conditions

Viscosity grade depends on operating temperature; multigrade oils are widely used in Europe to standardize products throughout the market and to avoid unacceptable seasonal oil changes.

16.6 HOUSINGS AND SEALS

The functions of housings are the following:

- Reacting to forces and torques applied by the contained parts and distributing the resultant forces to interfaces with the engine and power train suspension
- Maintaining the exact position of parts contained inside
- Wasting generated heat
- Insulating generated noise
- Allowing simple gearbox assembly and disassembly

Housing lay-out can be classified according to three alternative architectures:

- Through housing, when bearings seats are cut on the same housing element, with results that are particularly stiff and easily machined; openings closed by removable covers are available to allow assembly and disassembly of interior parts.
- End loaded housings; the housing is cut transversely to shafts in two halves. Therefore the bearing seats of the same shaft rest on different parts of the housing.
- Top loaded housings; these are cut along shafts in two halves, so that each bearing rests on two different half seats.

Also in these last cases, additional covers must be provided, to make assembly and disassembly possible.

If the housing is divided in two halves each part is machined separately during most of the cycle. Final boring of bearing seats will be made on assembled parts, to grant the necessary tolerance; there are therefore no clearance location pins that allow unambiguous half housing assembly.

The most widely used architecture is the second; it shows the advantage of easier assembly and, on industrial vehicle gearboxes, allows organization by modules (clutch, splitter, gearbox, reducer, accessories, etc.) assembled in different versions.

Housings are usually made of aluminium and, sometimes, magnesium for weight reduction; they show a large number of local reinforcements, such as ribs and webs, to achieve maximum stiffness with contained weight.

The example in Fig. 16.9 shows inclined ribs to increase torsional stiffness to the reaction forces of the shafts.

Housings must have breathers. As a matter of fact, lubricant does not occupy all available interior volume, to contain weight and friction losses; without

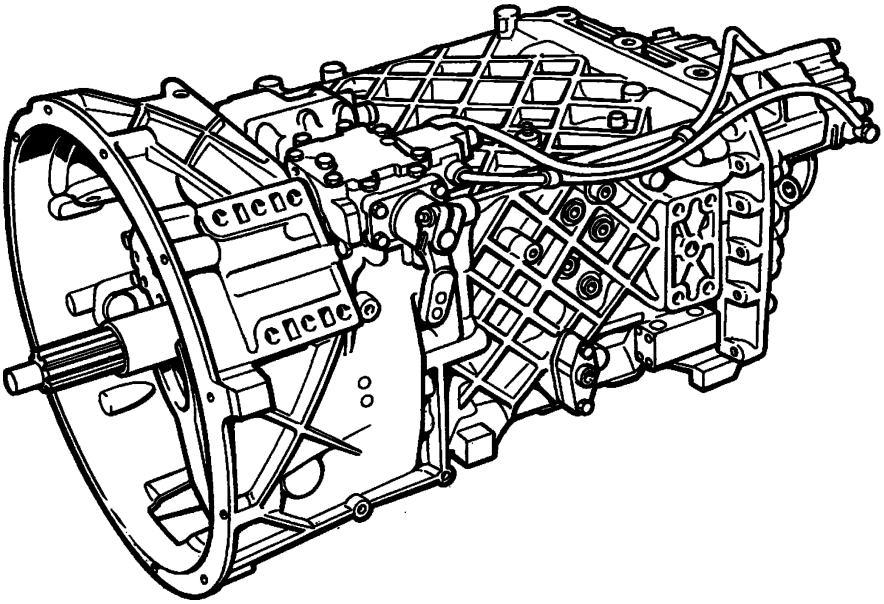


FIGURE 16.9. Industrial vehicle gearbox housing, characterized by a considerable number of inclined ribs; these increase global torsional stiffness and panel bending stiffness (Iveco).

breathers, the air in the free space would change pressure because of temperature variation, causing problems for seals.

Air must exit during vehicle operation and re-enter at stops; dust and other pollutants must be kept away. The breather is like a cap; it provides suitable openings with separation labyrinths and filters. The latter are made with low density sintered metal and are provided on the cap.

Rotary and sliding seals must be carefully designed. Seals must be completely tight; even small leakages are now unacceptable for reasons of environmental pollution and the need for consequent refills.

To seal fixed parts pre-formed gaskets are used (Fig. 16.10, middle), or in situ polymerized gaskets (Fig. 16.10, right). In this case, gaskets are made with synthetic materials that are distributed on parts as a paste and polymerize, becoming solid, after assembly; for this reason, covers must have suitable teeth (indicated by an arrow in the figure) to avoid intrusion of paste, after bolt tightening. Covers must also have suitable projections to make gasket rupture more easily at disassembly.

For limited dimension and round covers, O-rings are also used (Fig. 16.10, left). In the case of pre-formed gaskets the number of bolts and the cover plate stiffness must grant an almost constant pressure contact.

To verify this fact a photographic pressure sensitive film can be useful.

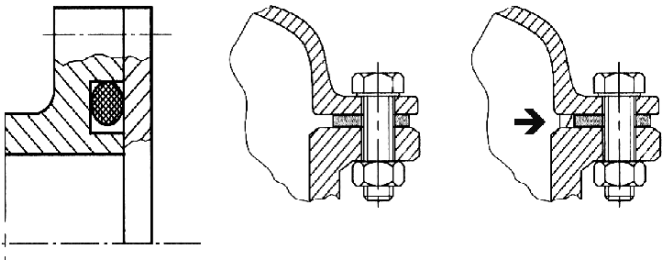


FIGURE 16.10. Examples of seals on covers; from the left, an O-ring seal for round covers, a pre-formed gasket and a gasket polymerized in situ.

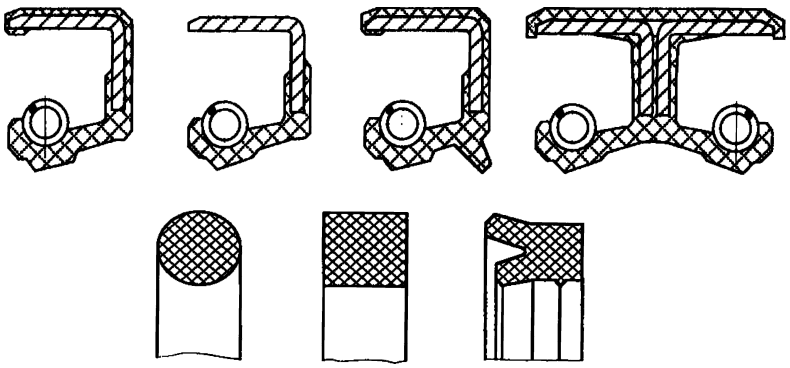


FIGURE 16.11. Examples of rotary seals (top) and sliding seals (bottom).

Rotary seals are of the lip type with a coil spring that clamps the rotary shaft (Fig. 16.11, top). Seals must be assembled with their springs inside the housings, to improve tightening for pressure increases due to temperature; a second lip can be added on the seal, to protect the sealing circle from dust contamination.

Sliding seals or small angle rotary seals (selector rods, Fig. 16.11, bottom) are made with O-ring or square rings.

Seals for multi-disc clutch actuation pistons are subject to high pressure and are made with rectangular rings with a pressure sensitive lip (Fig. 16.11, bottom right); these seals must be correctly oriented at assembly.

16.7 OUTLINE OF TEST TECHNOLOGIES

To verify the functions and reliability of a gearbox, suitable test activities are most conveniently performed after calculations are complete; these tests must be made on different prototypes, to provide a suitable level of confidence.

Test activities can be classified according to their time position, with reference to the development process of a new gearbox.

Tests will be run to demonstrate design adequacy, performed on a limited number of prototypes that are manufactured with experimental tools.

Following these tests will be a second series of tests, suitable for demonstrating the adequacy of the manufacturing process, performed on a significant number of prototypes manufactured with mass production tools.

These tests are performed on benches and prototype vehicles; they are followed by a vehicle reliability demonstration program that should confirm transmission reliability and identify residual problems.

The same test cycle must be separately performed on all supplied parts.

Test activities can also be classified according to expected results; from this point of view we identify *functional* and *reliability tests*.

The fundamental characteristic of functional tests is that they are performed in a short time, because sudden changes in result are not expected.

Some functional tests can be repeated on the same prototype at different times of its life. For example, mechanical efficiency measurements should be repeated at new and after run-in, to verify consequent improvements. In the same way, leakage tests must be performed on new gearboxes to verify the adequacy of the design and production process and at the end of useful life to detect unacceptable variations due to wear.

Typical functional test results are characteristic measurements that must be compared with project objectives.

Functional tests include the following:

- Lubrication tests, where it is verified that oil reaches all points to be lubricated, even when the gearbox is inclined in three directions, according to the vehicle mission
- Leakage test of lubricant oil
- Power absorption tests, to be performed at any possible input torque, engine speed and gearbox speed
- Selection and engagement forces, to be performed at different gearbox speed, vehicle speed and different meaningful oil temperatures
- Noise emission tests at idle and at different working conditions at different speeds
- Operating temperature measurements
- Misuse and abuse tests

All these tests can be performed on a small number of prototypes that must be machined and assembled with those dimensions relevant to phenomena under study, as close to tolerance limits as possible; for example rattle noise tests should be performed with the widest angular clearance allowed by the drawing specifications.

Endurance tests consist in having the component operate for the expected life according to different possible mission profiles. Expected failures are those that occur after useful life only; if they occur prematurely, they must be analyzed to design corrective countermeasures. Repetition of all scheduled tests is in any case requested, until success is obtained.

Reliability can be demonstrated by repeating endurance tests on a statistically significant number of prototypes.

Almost all functional and endurance tests can be performed on a bench or on a vehicle; it is useful to test a vehicle only for result confirmation, when success has been achieved in an adequate number of bench tests. The latter are, in fact, easier to supervise and failures are easier to analyze.

The stand-by time of a vehicle undergoing tests and the kind of damage consequent to a gearbox failure are unacceptable, considering the high cost of prototype fabrication.

Transmission test benches are particularly simple and include a foundation block on which a complete transmission can be installed. The transmission can be put in rotation by an actual engine or by an electric motor; this last is more appropriate for long endurance tests.

When an electric motor is used in particular tests, a control circuit is necessary to produce an input torque with the periodic irregularity of the internal combustion engine. The same result can be obtained by connecting motor and transmission through a torque pulsator.

According to the type of test, the bench schemes drawn in Fig. 16.12 may use a brake to simulate vehicle resistance.

Brakes can be coupled to variable inertia flywheels when vehicle inertia must be reproduced; brake and flywheel assembly can be replaced by a suitably controlled motor/generator, able to emulate vehicle resistance and recover part of the wasted energy.

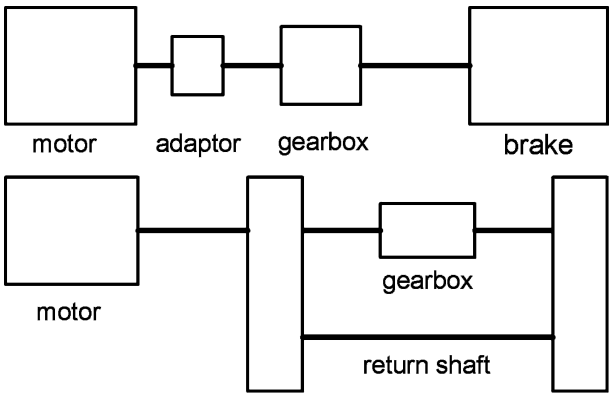


FIGURE 16.12. Schemes for transmission test benches; at the lower part of the figure is the scheme of a recirculation power bench.

FIGURE 16.13. Interior of a typical modern test cell for complete transmissions. This kind of test cell can be adapted to all kind of tests, including acoustic emission measurement. It is acoustically reverberant, but can be changed to an anechoic cell, by encapsulating the transmission only in a suitable cabin.

Sometimes (Fig. 16.12, at bottom), when tests imply constant input torque, shafts are connected through a transmission line. The transmission line is preloaded through a constant torque, which stresses the gearbox at the desired level. In this case motor power only compensates for friction resistance.

The picture in Fig. 16.13 shows the interior of a typical modern test cell for complete transmissions.

This kind of test cell can be adapted to all kind of tests, including acoustic emission measurement. It is acoustically reverberant, but can be changed to an anechoic cell, by encapsulating the only transmission in a suitable cabin.

We can see on the left the electric motor and its control system, the transmission test bed and the torsiometer shaft to measure output torque. In the small picture below, taken from the control console, we can see the electric brake.

The electric brake has a maximum power of 220 kW and a maximum torque of 600 Nm and can operate up to 7,000 rpm, simulating torque fluctuations of an internal combustion engine in the range of frequencies between 0 and 500 Hz.

The brake can absorb up to 200 kW, with a maximum torque of 3,000 Nm at 650 rpm; a maximum transmission ratio of 5 at maximum torque can be simulated, sufficient to test gearbox and final drive separately, for the maximum torque, or together for reduced torque.

REFERENCES OF VOLUME I

Part I

1. L. Baudry de Saunier, *L'automobile théorique et pratique*, Omnia, Paris, 1900
2. O. C. Schmidt, *Practical Treatise on Automobiles*, The American Textbook, Philadelphia, PA, 1909
3. W. Neubecker, *Antique Automobile Body Construction and Restoration*, Post Publication, Arcadia 1912
4. M. Peter, *Der Kraftwagen*, R. C. Schmidt, Berlin, 1937
5. M. Serruys, *La suspension et la direction des véhicules routiers*, Dunod, Paris, 1947
6. M. Boisseaux, *L'automobile, méthodes de calcul*, Dunod, Paris, 1948
7. J. C. Maroselli, *L'automobile et ses grands problèmes*, Larousse, Paris, 1958
8. M. G. Bekker, *Off-the-Road Locomotion*, University of Michigan Press, Ann Arbor, MI, 1960
9. J. P. Norbye, *Sports Car Suspension*, Sport Car Press, New York, 1965
10. J. Pawlowski, *Vehicle Body Engineering*, Business Books, London, 1969
11. G. Oliver, *Cars and Coachbuilding*, Sotheby Parke Bernet, London, 1981

12. I. S. Ageikin, *Off-the-Road Mobility of Automobiles*, Balkema, Rotterdam, 1987

13. D. Giacosa, *Progetti alla FIAT*, Automobilia, Torino, 1988

14. T. D. Gillespie, *Fundamentals of Vehicle Dynamics*, SAE, Warrendale, PA, 1992

15. M. Mitschke, *Dynamik der Kraftfahrzeuge*, Springer, Berlin, 1995

16. W. F. Milliken and D. L. Milliken, *Race Car Vehicle Dynamics*, SAE, Warrendale, PA, 1995

17. K. Newton *et al.*, *The Motor Vehicle*, SAE, Warrendale, PA, 1996

18. J. Reinpell, H. Stoll, *The Automotive Chassis*, Arnold, London, 1996

19. P. L. Bassignana *et al.*, *Storia fotografica dell'industria automobilistica italiana*, Boringhieri, Torino, 1998

20. J. Fenton, *Handbook of Automotive Body and Systems Design*, Professional Engineering Publishing, London, 1998

21. J. Fenton, *Handbook of Automotive Powertrain and Chassis Design*, Professional Engineering Publishing, London, 1998

22. H. Heisler, *Vehicle and Engine Technology*, Arnold, London, 1999

23. J. Fenton, *Handbook of Vehicle Design Analysis*, SAE, Warrendale, PA, 1999

24. J. Appian-Smith, *An Introduction to Modern Vehicle Design*, SAE, Warrendale, PA, 2002

25. J. Brown *et al.*, *Motor Vehicle Structures: Concepts and Fundamentals*, SAE, Warrendale, PA, 2002

Part II

1. L. Baudry de Saunier, *L'automobile théorique et pratique*, Omnia, Paris, 1900

2. O. C. Schmidt, *Practical Treatise on Automobiles*, The American Textbook, Philadelphia, PA, 1909

3. A. Seniga, *Il meccanismo di trasmissione negli automobili*, Biblioteca d'automobilismo e d'aviazione, Milano, 1912

4. E. B. Butler, *Transmission Gears*, Griffin, London, 1917

5. M. Peter, *Der Kraftwagen*, R. C. Schmidt, Berlin, 1937

6. M. Boisseaux, *L'automobile, méthodes de calcul*, Dunod, Paris, 1948

7. W. H. Crouse, *Automotive Transmissions and Power Trains*, McGraw-Hill, New York, 1955

8. P. Patin, *Les transmissions de puissance*, Eyrolles, Paris, 1956

9. D. Thirlby, *The Chain Driven Frazer Nash*, McDonald, London, 1965

10. G. Rogliatti, C. Valier, L. Giovanetti, *La frizione nel tempo*, Valeo, Torino 1980

11. D. Giacosa, *Progetti alla FIAT*, Automobilia, Torino, 1988

12. H. J. Schöpf, G. Jürgens, J. Pickard, *Das neue Fünfgang-automatikgetriebe von Mercedes-Benz*, ATZ, 91, 1989

13. Many Authors, *Design Practices: Passenger Car Automatic Transmissions,* SAE, Warrendale, 1994

14. J. Fenton, *Handbook of Automotive Powertrain and Chassis Design*, Professional Engineering Publishing, London, 1998

15. H. Heisler, *Vehicle and Engine Technology*, Arnold, London, 1999

16. G. Lechner, H. Naunheimer, *Automotive Transmissions, Fundamentals, Selection, Design and Application*, Springer, Berlin, 1999

17. R. K. Jurgen et al., *Electronic Transmission Control*, SAE, Warrendale, PA, 2000

18. J. Happian-Smith, *An Introduction to Modern Vehicle Technology*, SAE, Warrendale, PA, 2002

19. J. Greiner et al., *7-Speed Automatic Transmission from Mercedes-Benz*, ATZ, 105, 2003

INDEX

4WS
    Four Wheel Steering, *see* RWS

ABS
    Anti-lock Braking System, 326
accelerometer, 336
Ackermann, 10
Ackermann angle, 242
adhesion (rubber), 59
airbag, 263
Aisin, 386
Alfa Romeo, 27, 170, 172, 189
aligning
    coefficient, 113
    torque, 101
Allison, 386
Anderson, 395
angle
    Ackermann -, 242
    attitude -, 240
    camber -, 59, 87, 137
    caster -, 217, 263
    crown -, 55
    inclination -, 59
    king-pin -, 137, 263
    steering -, 242
    tire sideslip -, 59, 64, 86, 100
    toe -, 138
    vehicle sideslip -, 240
annulus, 555
anti-lock devices, *see* ABS
anti-roll bar, 161
anti-spin devices, *see* ASR
apparent density, 65
aquaplaning, 61, 97, 104
artillery wheel, 36
aspect ratio (tire), 41, 56
ASR
    Anti-spin Regulator, 333
attitude angle, 240
Audi, 434, 507, 513, 552, 554, 563, 572
auxiliary chassis structure, 351
auxiliary frame, 356

B10, *see* endurance
bar mechanism, 456
BAS
    Brake Assist System, 334
beam model, 375

bearing
first generation -, 158
second generation -, 158
third generation -, 158
bedding number, 69
Benz, 12, 395
BMW, 324
Bodmer, 413
Bollée, 11
Bordino, 10
Borg Warner, 489
Bosch, 326, 330
brake
band -, 42
disc -, 275
distribuor, 284
drum -, 42, 280
efficiency, 303
external shoe -, 42
parking -, 272, 275
braking
circuit, 274
fluid, 283
brush model, 63
bulldozing resistance, 73

cable mechanism, 456
efficiency, 473
caliper, 276
caliper interlock, 452
camber
angle, 37, 59, 87, 137
recovery, 138
stiffness, 111
stiffness coefficient, 111
thrust, 106
carcass, 56
carpet plot, 104
carrier, 555
caster angle, 137, 217, 263
characteristic curves, 482
chassis, xv
Citroën, 348
clutch
band -, 409
disc -, 410
electromagnetic -, 415, 473
lock-up -, 487
pulled spring -, 467
pushed spring -, 467
shoe -, 410
tapered disc -, 408
cohesive soil, 66
cohesivity, 71
compressor, 293
conicity force, 107
constant gear, 432
constant velocity joint, 537
contact
patch, 62
pressure, 88
control valve, 294
conventional ply tires, 56
cornering
force, 100
stiffness, 110
Cotal, 415
countershaft, 401
crash, 215, 263
critical speed, 79
cross member, 47
cross ply tires, 56
crown angle, 55
Cugnot, 394
curvature gain, 246
CVT
Continuosly Variable Transmission, 546

DAF, 422
Daimler, 12, 395
damping
characteristic, 168
coefficient, 168
Dana, 386
De Dion and Bouton, 412
De Dion axle, 198
De Rivaz, 395
Dejbjerg, 36
Delphi, 321

design point, 481
diaphragm spring, 462, 464
differential, 505
    controlled -, 519
    front drive -, 507
    locking -, 509
    mechanical efficiency, 513
    rear drive -, 506, 507
    self-locking -, 519
direct drive, 432, 438
disc, 54
disc brake, 275
disengagement
    load, 473
    mechanism, 463, 464
    stroke, 473
distributor valve, 295
Dodge, 418
double clutching, 400, 546
double shock, 496
double tube
    shock absorber, 165
driven plate, 461, 467
driving plate, 461
drum brake, 280
Dubonnet, 22
Duesemberg, 44
Dunlop, 39
Dynaflow, 418

Eaton, 386
EBD
    Electronic Brake Distributor, 331
effective rolling radius, 74, 90
EHB
    Electronic Hydraulic Brakes, 339
elastic bushing, 153
elasticity (rubber), 59
elasto-kinematic behavior, 138, 217, 263
Elliot, 10
EMB
    Electro Mechanical Brakes, 340
emergency braking system, 272
endurance, 594, 597
engine brake, 575
EPB
    Electronic Parking Brake, 339
EPS
    Electric Power Steering, 259
equator plane, 55
Erech, 33
ESP
    Electronic Stability Program, 332
external shifting mechanism, 449

Föttinger, 416, 475
fading, 274, 283
Falchetto, 20
fatigue, 212, 262, 378, 597
Ferguson, 522
Ferodo, 42, 407
FIAT, 13, 22, 25, 30, 46, 48, 50, 141, 148, 178, 182, 183, 190, 195, 319, 323, 342, 352, 386, 397, 399, 406, 407, 435, 450–453, 456, 489, 491, 507, 511, 606
final drive, 391, 505, 509
finger, 450
fixed caliper, 277
flange, 54
floating bar, 179
Ford, 14, 25, 186, 413
fork, 450
four wheels drive, 389
Frazer Nash, 401
friction
    angle, 71
    circle, 119
    coefficient, 61
front wheel drive, 387
Frood, 42, 407
fuel consumption map, 567

gearbox
automated -, 544
continuously variable -, 546
countershaft -, 406, 425
dog clutch -, 401
double stage -, 388, 425
dual-clutch -, 552
epicycloidal -, 413, 415
full automatic -, 544
Fuller -, 442, 447
industrial vehicle -, 437
leather belt -, 395
multi-stage -, 425
range changer -, 391, 440
robotized -, 544
rubber belt -, 422
schemes, 425
semi-automatic -, 544
single stage -, 387, 397, 425
sliding trains -, 399
splitter -, 391, 440, 446
stepped -, 546
Getrag, 386, 433
globoidal screw, 252
GM, 386, 418
Goodyear, 40
Gough diagram, 103
grating, 603
Graziano, 386
Griffith, 399
guided vehicles, 239
Gyromatic, 418
gyrometer, 336

half shaft, 533, 537
HH
Hill Holding, 339
Honda, 26, 319, 321, 325, 552, 569
Hooke, 536
hump, 54
Hydramatic, 416
hydraulic
clutch, 475
losses, 480
hydroforming, 357
hypoid gear, 507

idler, 397, 426
impact loss, 480
inclination angle, 59
inflation pressure, 75, 84
integral chassis structure, 351
internal shifting mechanism, 449
Isotta Fraschini, 43
Iveco, 201, 202, 205, 206, 446, 447, 509, 534, 539, 577, 610

Jatco, 386
Jeantaud, 11, 243
jerk, 544
jump-in pressure, 289

kinematic behavior, 138
kinematic steering, 241
king-pin, 10
angle, 137, 263
axis, 137
offset, 137, 217

Längensberger, 10
ladder structure, 46
Lancia, 20, 29, 50, 177, 357, 359
lateral-longitudinal interaction, 118
leading shoe, 303
leaf spring, 14, 195
(cantilever), 14
liquid limit, 66
load factor, 56
load profile, 602
locking coefficient, 519
longitudinal
force, 90
force coefficient, 93
slip, 90
slip speed, 63
trail, 137, 217
low-speed steering, 241
lower arm, 149

lubrication
boundary -, 609
hydrodynamic -, 609
mixed -, 609
LuK, 569, 572

magic formula, 98, 115
Marcus, 399
Marelli, 547, 549, 550
master pump, 282
McPherson, 25
mechanical
efficiency, 428
efficiency map, 429
losses, 428
Mercedes, 31, 175, 188, 506, 511, 561–563
Michelin, 39–41
Miner, 214, 602
misuse, 214
mono tube
shock absorber, 165
MSR
Motor Spin Regulator, 334
mule car, 231

New Venture Gear, 386
Newton, 523

overdrive, 438

Pacejka, 121, 326
pad, 276
Panhard bar, 198
Pecqueur, 399
performance coefficient, 481
pillar, 50, 354
piloted vehicles, 239
pitting, 597
planetary wheel, 506
plastic limit, 66
plasticity
index, 66
pressure, 70
platform, xvi, 49
plunger interlock, 451
ply-steer force, 108
pneumatic trail, 101
Porsche, 23
positive engagement, 495
positivity, 455
power brake gain, 289
power steering, 255
powershift, 545
powertrain, 388
pre-synchronization, 495
precision, 455
pressure plate, 463, 464
propeller shaft, 533, 534

radial ply tires, 56
range, 546
rattle, 603
rattling, 434
Ravigneaux, 559, 561, 563
reactor, 476
real density, 65
rear engine drive, 388
recirculating ball
steering box, 253
reducer, 511
relative content of water, 66
relaxation length, 127
reliability, 597
retarder, 291, 575
rim diameter, 54
roll
axis, 221
steer, 221
rolling
coefficient, 78
radius, 74
resistance, 75
Rolls-Royce, 18, 44
Rudge-Whitworth, 38
RWS
Rear Wheel Steering, 318
angle dependent, 319
dynamic speed dependent, 321

model following, 322
speed dependent, 321
Rzeppa, 537, 540

S-cam, 550
S-deformation, 17
sandy soil, 66
Sankey, 38
satellite, 506, 555
saturation load, 287
saturation pressure, 290
scuffing, 598
Selden, 399
selection plane, 431
self-adjusting mechanism, 471
separated chassis structure, 351
service braking system, 272
shift
with power interruption, 545
without power interruption, 545
shimmy, 18, 257
shock absorber, 163
(magneto-rheologic), 346
side beam, 46
side force coefficient, 102
sideslip angle (tire), 59, 64, 86, 100
sideslip angle (vehicle), 240
Sizair Naudin, 403
sky-hook damping, 345
sleeve, 450
slick tire, 58
sliding caliper, 277
slip
stiffness, 93
velocity, 91
slip (torque converter), 474
smoothness, 455
snowy soil, 67
soil stiffness, 69
speed
transmission ratio, 474
triangle, 478
spherical joint, 156
spoked wheel, 38
sprung mass, 135
static margin, 511
steer by wire, 325
steering
angle, 242
error, 217, 242
knuckle, 156
Stephens, 19
stick and slip, 456
stop spring, 159
stopping distance, 273
structural surface, 372
strut, 157
Studebaker, 24
subframe, 356
sun, 555
suspension
active -, 135, 341
adaptive -, 341
anti-dive -, 227, 350
anti-lift -, 227
anti-squat -, 227, 350
controlled -, 341
dependent -, 134
double wishbone -, 171
flexibility, 218
four-link -, 222
full active -, 350
guided trailing arm -, 185
independent -, 134, 139
Mac Pherson -, 141
multilink -, 139, 188
passive -, 135, 340
rigid axle -, 195
roll center, 221
semi-independent -, 134
semi-trailing arm -, 182
SLA -, 140
slow active -, 350
stroke, 136
swing arm -, 226
three-link suspension, 223
trailing arm -, 178, 226
twist beam -, 190
virtual center -, 175

swampy soil, 67
swinging shackle, 14, 195
synchronization, 495
synchronizer
    hub, 489
    multiple -, 490
    ring, 490
    simple -, 489

Thomson, 39
thrust bearing, 470
tip-in load, 287, 289
tire
    damping, 89
    stiffness, 90, 125
    temperature, 83
    testing machines, 129
    vibration, 79, 125
toe angle, 138
Torotrak, 574
torque
    steering, 189
    transmission ratio, 474
Torsen, 513, 521
torsion damper, 462, 468
track, 138
traction
    coefficient, 94, 103
traction fluids, 572
trailing shoe, 303
transfer box, 391, 505, 510
transmissibility, 125
transmission efficiency, 474
transmission ratio
    speed -, 474
    torque -, 474
transmittable torque, 461
transmitted torque, 461
tread, 57
Trilok, 416, 483
tripod joint, 542
trunnion cross, 536
tube tire, 55
tubeless tire, 55
Turicum, 404

universal joint, 536
unsprung mass, 135

vacuum booster, 287
Valeo, 462, 464, 468–471
Van Doorne, 569, 570
vapor-lock, 283
VDC
    Vehicle Dynamics Control, 332
vehicle trim, 134
Volkswagen, 23, 30, 49
VSC
    Vehicle Stability Control, 332

Wöhler, 214, 601
Watt, 29, 395
Watt quadrangle, 198
wear, 598
wheel
    center of rotation, 90
    reference system, 58
    speed sensor, 335
wheelbase, 138
whine, 603
whistle, 603
Wilson, 414, 557

yoke, 536

ZF, 343, 386, 520